356
1.9.70

Vol. 43, No. 2 March, 1946

Psychological Bulletin

EDITED BY
JOHN E. ANDERSON, University of Minnesota

WITH THE CO-OPERATION OF
S. H. BRITT, McCann-Erickson, Inc., New York; W. T. HERON, University of Minnesota; W. A. HUNT, Northwestern University; J. G. JENKINS, University of Maryland; D. G. MARQUIS, University of Michigan; A. W. MELTON, University of Missouri; J. T. METCALF, University of Vermont.

88
4782

CONTENTS

General Reviews and Summaries:

The Study of Eye Movements in Reading: Miles A. Tinker, 93.

Color Adaptation to 1945: Jozef Cohen, 121.

The Effects of Noise: F. K. Berrien, 141.

Studies in Time Perception: A. R. Gilliland, J. Hofeld, G. Eckstrand, 162.

Book Reviews: 177.

PUBLISHED BI-MONTHLY
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second-class mail matter at the post office at Evanston, Illinois, under the Act of March [illegible], 1879. Additional entry at the post office at Menasha, Wisconsin. Application has been made for the transfer of the original entry from the post office at Evanston, Illinois, to the post office at Washington, D.C.

[illegible] Edn. [illegible]
[illegible]
7. 9. 70
No. J - 356

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

Vol. 43, No. 3 May, 1946

Psychological Bulletin

EDITED BY

JOHN E. ANDERSON, UNIVERSITY OF MINNESOTA

WITH THE CO-OPERATION OF

S. H. BRITT, MCCANN-ERICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTHWESTERN UNIVERSITY; J. G. JENKINS, UNIVERSITY OF MARYLAND; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, UNIVERSITY OF MISSOURI; J. T. METCALF, UNIVERSITY OF VERMONT.

CONTENTS

General Reviews and Summaries:

Language and Psycholinguistics: A Review: N. H. PRONKO, 189.

Objective Measurement of Clinical Status in Psychopathological Research: DANIEL I. MALAMUD, 240.

History of Department of Psychology at University of Chicago: FORREST A. KINGSBURY, 259.

New Statistical Criteria for Learning and Problem Solution in Experiments Involving Repeated Trials: DAVID A. GRANT, 272.

Book Reviews: 283.

Books and Materials Received: 288.

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1227 Nineteenth Street, N.W., Washington 6, D.C.

Subscription price, $7.00 per year
Entered as second class mail matter at the post office at Washington, D.C., under the Act of March 3, 1879.
Additional entry at the post office at Menasha, Wisconsin.

PUBLICATIONS OF
The American Psychological Association, Inc.

PSYCHOLOGICAL REVIEW
HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions of a theoretical and integrative nature; bi-monthly.
Subscription: $5.50 (Foreign $5.75). Single copies, $1.00

PSYCHOLOGICAL BULLETIN
JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, critical and analytical summaries of psychological fields or subject matter; bi-monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

PSYCHOLOGICAL MONOGRAPHS
JOHN F. DASHIELL, *Editor*
University of North Carolina

Contains longer researches and laboratory studies which appear as units; published at irregular intervals.
Subscription: $6.00 per volume of about 350 pages (Foreign $6.30).
Single copies, price varies according to size.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY
GORDON W. ALLPORT, *Editor*
Harvard University

Contains original contributions in the field of abnormal and social psychology, reviews and notes; quarterly.
Subscription: $5.00 (Foreign $5.25). Single copies, $1.50.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY
SAMUEL W. FERNBERGER, *Editor*
University of Pennsylvania

Contains original contributions of an experimental character; bi-monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

PSYCHOLOGICAL ABSTRACTS
WALTER S. HUNTER, *Editor*
Brown University

Contains non-critical abstracts of the world's literature in psychology and related subjects; monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $.75

JOURNAL OF APPLIED PSYCHOLOGY
DONALD G. PATERSON, *Editor*
University of Minnesota

Contains material covering the application of psychology in business, industry, education, etc.; bi-monthly.
Subscription: $6.00 (Foreign $6.50). Single copies, $1.25.

APPLIED PSYCHOLOGY MONOGRAPHS
HERBERT S. CONRAD, *Editor*
College Entrance Examination Board, Princeton

Contains longer researches and studies in the fields of applied psychology; published at irregular intervals.
Single copies only, price varies according to size.

JOURNAL OF CONSULTING PSYCHOLOGY
MRS. J. P. SYMONDS, *Editor*
525 West 120 Street, New York 27, N.Y.

Contains articles in the field of clinical and consulting psychology and individual guidance; bi-monthly.
Subscription: $3.00. Single copes, $.60.

THE AMERICAN PSYCHOLOGIST
DAEL L. WOLFLE, *Editor*
1227 Nineteenth Street N.W., Washington 6, D.C.

The professional journal of The American Psychological Association, Inc.; contains articles on trends in psychology, proceedings of the Association and its affiliated organizations and divisions, psychological news and notes, and special announcements; monthly.
Subscription: $7.00 (Foreign $7.50). Single copies, $.75.

Subscriptions, orders, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
1227 NINETEENTH STREET N.W., WASHINGTON 6, D.C.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

Vol. 43, No. 4 July, 1946

Psychological Bulletin

EDITED BY

JOHN E. ANDERSON, UNIVERSITY OF MINNESOTA

WITH THE CO-OPERATION OF

S. H. BRITT, McCANN-ERICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTHWESTERN UNIVERSITY; J. G. JENKINS, UNIVERSITY OF MARYLAND; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, UNIVERSITY OF MISSOURI; J. T. METCALF, UNIVERSITY OF VERMONT.

CONTENTS

General Review and Summary:

Opinion-Attitude Methodology: QUINN McNEMAR, 289.

Note:

Reply to Elder's Note on Dunlap's Remedy for Color Vision: KNIGHT DUNLAP, 375

Book Reviews: 376.

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Massachusetts & Nebraska Aves., N.W., Washington 16, D.C.

Subscription price, $7.00 per year
Entered as second class mail matter at the post office at Washington, D.C., under the Act of March 3, 1879.
Additional entry at the post office at Menasha, Wisconsin.

PUBLICATIONS OF

The American Psychological Association, Inc.

Psychological Review

Herbert S. Langfeld, *Editor*
Princeton University

Contains original contributions of a theoretical and integrative nature; bi-monthly.
Subscription: $5.50 (Foreign $5.75). Single copies, $1.00

Psychological Bulletin

John E. Anderson, *Editor*
University of Minnesota

Contains critical reviews of books and articles, critical and analytical summaries of psychological fields or subject matter; bi-monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

Psychological Monographs

John F. Dashiell, *Editor*
University of North Carolina

Contains longer researches and laboratory studies which appear as units; published at irregular intervals.
Subscription: $6.00 per volume of about 350 pages (Foreign $6.30).
Single copies, price varies according to size.

Journal of Abnormal and Social Psychology

Gordon W. Allport, *Editor*
Harvard University

Contains original contributions in the field of abnormal and social psychology, reviews and notes; quarterly.
Subscription: $5.00 (Foreign $5.25). Single copies, $1.50.

Journal of Experimental Psychology

Samuel W. Fernberger, *Editor*
University of Pennsylvania

Contains original contributions of an experimental character; bi-monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

Psychological Abstracts

Walter S. Hunter, *Editor*
Brown University

Contains non-critical abstracts of the world's literature in psychology and related subjects; monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $.75.

Journal of Applied Psychology

Donald G. Paterson, *Editor*
University of Minnesota

Contains material covering the application of psychology in business, industry, education, etc.; bi-monthly.
Subscription: $6.00 (Foreign $6.50). Single copies, $1.25.

Applied Psychology Monographs

Herbert S. Conrad, *Editor*
College Entrance Examination Board, Princeton

Contains longer researches and studies in the fields of applied psychology; published at irregular intervals.
Single copies only, price varies according to size.

Journal of Consulting Psychology

Mrs. J. P. Symonds, *Editor*
525 West 120 Street, New York 27, N.Y.

Contains articles in the field of clinical and consulting psychology and individual guidance; bi-monthly.
Subscription: $3.00. Single copies, $.60.

The American Psychologist

Dael L. Wolfle, *Editor*
1227 Nineteenth Street N.W., Washington 6, D.C.

The professional journal of The American Psychological Association, Inc.; contains articles on trends in psychology, proceedings of the Association and its affiliated organizations and divisions, psychological news and notes, and special announcements; monthly.
Subscription: $7.00 (Foreign $7.50). Single copies, $.75.

Subscriptions, orders, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MASSACHUSETTS AND NEBRASKA AVES., N.W., WASHINGTON 16, D.C.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

Vol. 43, No. 5 September, 1946

Psychological Bulletin

EDITED BY
JOHN E. ANDERSON, UNIVERSITY OF MINNESOTA

478288

WITH THE CO-OPERATION OF
S. H. BRITT, MCCANN-ERICKSON, INC., NEW YORK; D. A. GRANT, UNIVERSITY OF WISCONSIN; W. T. HERON, UNIVERSITY OF MINNESOTA; W. A. HUNT, NORTHWESTERN UNIVERSITY; J. G. JENKINS, UNIVERSITY OF MARYLAND; D. G. MARQUIS, UNIVERSITY OF MICHIGAN; A. W. MELTON, UNIVERSITY OF MISSOURI; J. T. METCALF, UNIVERSITY OF VERMONT.

CONTENTS

General Reviews and Summaries:

The Validity of Personality Questionnaires: ALBERT ELLIS, 385.

The Present Status of Psychology in South America: MARGARET E. HALL, 441.

Book Reviews: 477.

Books and Materials Received: 487.

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Massachusetts & Nebraska Aves., N.W., Washington 16, D.C.

Subscription price, $7.00 per year
Entered as second class mail matter at the post office at Washington, D.C., under the Act of March 3, 1879.
Additional entry at the post office at Menasha, Wisconsin.

PUBLICATIONS OF

The American Psychological Association, Inc.

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions of a theoretical and integrative nature; bi-monthly.
Subscription: $5.50 (Foreign $5.75). Single copies, $1.00

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, critical and analytical summaries of psychological fields or subject matter; bi-monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Contains longer researches and laboratory studies which appear as units; published at irregular intervals.
Subscription: $6.00 per volume of about 350 pages (Foreign $6.30).
Single copies, price varies according to size.

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Contains original contributions in the field of abnormal and social psychology, reviews and notes; quarterly.
Subscription: $5.00 (Foreign $5.25). Single copies, $1.50.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

SAMUEL W. FERNBERGER, *Editor*
University of Pennsylvania

Contains original contributions of an experimental character; bi-monthly
Subscription: $7.00 (Foreign $7.25). Single copies, $1.25

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Contains non-critical abstracts of the world's literature in psychology and related subjects; monthly.
Subscription: $7.00 (Foreign $7.25). Single copies, $.75.

JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
University of Minnesota

Contains material covering the application of psychology in business, industry, education, etc.; bi-monthly.
Subscription: $6.00 (Foreign $6.50). Single copies, $1.25.

APPLIED PSYCHOLOGY MONOGRAPHS

HERBERT S. CONRAD, *Editor*
College Entrance Examination Board, Princeton

Contains longer researches and studies in the fields of applied psychology; published at irregular intervals.
Single copies only, price varies according to size.

JOURNAL OF CONSULTING PSYCHOLOGY

MRS. J. P. SYMONDS, *Editor*
525 West 120 Street, New York 27, N.Y.

Contains articles in the field of clinical and consulting psychology and individual guidance; bi-monthly.
Subscription: $3.00. Single copies, $.60.

THE AMERICAN PSYCHOLOGIST

DAEL L. WOLFLE, *Editor*
1227 Nineteenth Street N.W., Washington 6, D.C.

The professional journal of The American Psychological Association, Inc.; contains articles on trends in psychology, proceedings of the Association and its affiliated organizations and divisions, psychological news and notes, and special announcements; monthly.
Subscription: $7.00 (Foreign $7.50). Single copies, $.75.

Subscriptions, orders, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.
MASSACHUSETTS AND NEBRASKA AVES., N.W., WASHINGTON 16, D.C.

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN

Vol. 43, No. 2 March, 1946

Psychological Bulletin

THE STUDY OF EYE MOVEMENTS IN READING

MILES A. TINKER
University of Minnesota

The study of eye movements continues to be an important and much used technique for investigating the reading process. An overall summary of the field up to 1935 is given by Tinker (**108**). The last review (**107**) to appear in this journal was published in 1936. In the present review, material appearing during the years of 1935 through 1944 is considered. Bibliographies, critical evaluations, summaries of parts of the field, and textbook discussions will be found in references (**13, 37, 38, 41, 74, 95, 97, 103, 116, 117**). A few titles not available to the reviewer have been included in the bibliography for the sake of completeness (**6, 29, 30, 46, 47, 48, 49, 126**). Certain material on eye movements will not be reviewed since it is not closely related to the reading process. This includes reports on eye movements in viewing pictures and advertisements, certain studies of visual fixation, eye movements as related to visual attention, and the like.

The studies to be reviewed here group themselves as follows: techniques of measurement, analysis of the reading process, oculomotor coordination, factors in variability, special reading situations, heredity and certain environmental factors, eye movements in the reading clinic, training eye movements, and central versus peripheral processes.

TECHNIQUES OF MEASUREMENT

Apparatus. Although some of the new apparatus for recording eye movements consists of modifications of earlier techniques, others involve new principles. Brandt (**9**) describes an adaptation of the Dodge corneal reflection type of camera. Its special feature is simultaneous recording of the vertical and horizontal components of eye movements. A portable model of the Dodge type of camera, the Ophthalm-o-graph, is described by Taylor (**105**). Records of horizontal eye movements of both eyes are recorded simultaneously. Tiffin and Fairbanks (**106**) have developed a camera for simultaneously recording eye movements and voice reactions in oral reading. Miles' (**70**) note on the first eye-movement record photographed by Dodge has historical interest. A demonstration film on eye movements in reading has been prepared by Valentine, Troyer and Brown (**118**).

An electrical method of recording eye movements is finding some use. Changes in the corneo-retinal potentials are closely proportional to the sine of the angle of rotation of the eye. With proper placement of electrodes and with an amplifying system, the movements may be recorded mechanically or photographically. Miles (**71**) and Halstead (**40**) suggest that eye movements in a variety of situations may be measured by this technique. Its validity for recording eye movements in reading has been checked by Hoffman, Wellman and Carmichael (**43**). Electrical and corneal reflection records were made simultaneously while their subjects read and while they made horizontal and vertical eye movements in viewing a dot pattern. The only data given for the reading responses was a sample record. Although the two techniques yielded similar records, they were far from identical. Where extent of eye movements and duration of fixation pauses for viewing the dot patterns were measured, the agreement as shown by correlations is close but again the records do not yield identical results. Although this electrical technique has sufficient validity to justify its use in certain situations where it is impossible or inconvenient to photograph the corneal reflection, it is clear to the reviewer that it is not satisfactory in its present form for studying eye movements in reading.

Another electrical technique, recording photographically (electromyograms) the action currents generated by movement of extrinsic eye muscles during reading, has been used by Luckiesh (**61**) and Luckiesh and Moss (**62, 63, 64, 65**). The sample record given by Luckiesh (**61**, p. 128) is either not typical or the author has read the time intervals wrongly. He gives 0.120 sec. for time of back sweep when the typical value is approximately 0.050 sec. The duration of his interfixation movement also appears unduly long. Furthermore, the demarkation between successive fixation pauses is not always clear. A direct comparison of this technique with the corneal reflection method is needed to check the validity of the former. The corneal reflection technique of photographically recording eye movements, therefore, remains the most satisfactory method for use in reading investigations.

Reliability. The reliability or consistency of eye-movement records has been considered by several workers. Imus, Rothney and Bear (**44**), using fifty word samples of text, found that eye-movement measures had low reliability (r = .59 to .72). In another report Imus, Rothney and Bear (**45**) re-emphasize the unreliability of eye-movement records.

Employing the same material, i.e., that furnished with the Ophthalm-o-graph, Anderson (**3**), and Broom (**12**) note similar trends in low reliability. In a more extensive study Tinker (**112**) found that reliability of eye-movement measures was low only when short samples of text (5 to 15 lines) were employed. When 20 to 40 lines were read the reliabilities were about .80. Thus eye-movement records will have adequate reliability only when a satisfactorily long sample of reading is used. When only group comparisons are involved, short samples are adequate. But in the clinic, where individual diagnosis occurs, records from reading at least 20 lines of text are indicated. An even longer sample is desirable.

Validity. If eye-movement measures are to have meaning for validity evaluation there must be adequate comprehension of the material read. Users of the Ophthalm-o-graph for photographing eye movements have commonly employed short samples of reading material with questions on comprehension furnished with the apparatus. Imus, Rothney and Bear (**45**) have found this check on comprehension to be useless. Authors such as Broom (**12**), Anderson (**1, 3**) and Imus, Rothney and Bear (**44, 45**), have reported very low validity coefficients for eye-movement measures of reading when correlating eye-movement measures with scores on reading tests. Such a comparison constitutes an unsatisfactory test of validity since it is well known that reading scores derived from different reading situations are not related or yield only low correlations. Tinker (**112**) has shown that when the material read before the camera is strictly comparable to that in the performance tests the validity of fixation frequency and perception time (fixation frequency times pause duration) was very high ($r = -.80$ to $-.99$). Pause duration by itself was found not to be a valid measure of reading performance.

Other evaluation. Is a reader able to give a typical reading performance in the apparently artificial laboratory situation where the photographing of eye movements takes place? Tinker (**112**) has checked this and found that the performance was the same before the camera as when working at a table. This finding has been confirmed by Gilbert and Gilbert (**34**). It should be noted, however, that it is important to adapt a reader to the laboratory situation when photographing eye movements.

Stromberg (**101**) has questioned the common practice of plotting the fixation points of the eyes to show the location of these fixations on the material read. This has ordinarily been done for a single eye, assuming that the two eyes coordinate in fixating on the text. When Stromberg plotted the fixations for both eyes he found what appear to be startling discrepancies in binocular fixation. That is, the fixation of one eye frequently did not coincide with that of the other eye. Undoubtedly there has been unwarranted optimism concerning the usefulness of plotted fixations to show exactly where the eyes paused. Since one fixates an area rather than a point, and because movement of the reflected beam does not correspond exactly to the visual angle of movement, plotted fixations can only represent approximations to the correct place of fixation. One suspects that part of the binocular fixation discrepancy found by Stromberg was due to the reflected spot of light not being in corresponding locations on the two eyeballs, a condition which would prevent correspondence on the photographic record. Nevertheless, this study does raise an important question concerning interpretation of plotted fixations.

In general the evidence indicates that under adequate experimental conditions, eye-movement records are reliable and valid measures of reading performance.

ANALYSIS OF THE READING PROCESS

The fixational pause. Perception occurs during fixational pauses. Arnold and Tinker (5) made an extensive analysis of the fixational pause of the eyes. On the average a pause of .157 sec. was required to identify a letter and .172 sec. to accurately fixate a dot. These are longer than the .100 sec. required for a well cleared-up perception. In a variety of reading situations the mean pause duration varied from .217 to .404 sec. These are much longer than the pauses made in simple fixations requiring a minimum of perception. The longer durations in reading may be explained in terms of the requirements for comprehension and assimilation of the materials read. Luckiesch and Moss' (65) statement that reading pauses average about .150 sec. is a misrepresentation. Reading with comprehension cannot have a mean pause duration as short as this.

An excellent analysis of regressive pauses, i.e., pauses following backward eye movements within a line of print, in the reading of 9th and 10th grade children is given by Bayle (7). She found that regressive pauses tended to fall into six patterns: adjustment after the first fixation in a line, adjustment within a line when the span of vision is overreached in a too long forward move, regressions for verification, regressions during word analysis, regressions for phrase analysis, and regressions for re-examination of a whole line. It was found that regressions were caused by need to complete perception and by difficulties of interpretation due to failure to recognize the meaning of a word or failure to connect meaning of a word with the context. Bayle points out that regressive pauses are necessary parts of the reading process in analytical reading. Buswell (14) gives a cautious approval of this view but many writers ignore it. Recognition of the fact that regressions are legitimate and essential in certain kinds of reading would avoid some of the misconceptions involved in training eye movements (see p. 110).

The fixation span. This is the amount of material read in one fixation. The larger the span, the fewer the fixations and consequently the faster the reading. By using reading material with spaces between phrases and encouraging readers to fixate each phrase only once, Robinson (87) obtained the relation between perceptual span from tachistoscopically exposed material with fixation span in reading. For tachistoscopic material not in context this correlation was .52, for material in context (using a special tachistoscope) the correlation was .66. He concluded correctly that span derived from short exposures predicts eye-movement behavior inadequately. La Grone (55) and Knehr (51) also found a low correlation between perceptual span and fixation span. Dearborn and Anderson (27) have perfected a highly flexible and effective motion picture technique for exposing successive phrases in context to increase the fixation span.

In a study of adult reading, Buswell (14) emphasizes that fixation span or span of recognition is an important index of reading maturity, i.e., a narrow span indicates immature reading habits. He found no relation between visual efficiency and span. There is a similar emphasis upon fixation span in the author's second monograph (15). In all of Buswell's work (14, 15) there is a sound emphasis upon training to increase the span of recognition which directs attention to perception rather than to the motor aspects of eye movements.

Luckiesh and Moss (62) have studied the influence of typographical factors on extent of the fixation span. They found that span decreased slightly in going from four to 10 point type and that it increased slightly in going from 13 to 21 and 29 pica line widths. Their interpretations reveal a decided lack of understanding of the reading process. No account is taken of possible accompanying changes in either pause duration or efficiency of peripheral vision.

OCULOMOTOR COORDINATION

Rhythm. Much has been made of rhythm in eye movements. Buswell (14) states that "a mature reader covers a line of print with regular, rhythmic eye movements" but admits that reading habits should be flexible and that in some situations regressions are desirable. Taylor (105) has devised techniques for developing "rhythmical, left-to-right eye-movements." In the non-experimental literature the importance of rhythmical eye movements is stressed, i.e., regular forward movements along a line of print and about the same number of fixations from line to line are considered highly desirable. This emphasis is derived from the notion that the efficient reader is efficient because he is able to establish rhythmical eye-movement habits more readily and that favorable typography fosters rhythmical movements. Sisson (94) has made a quantitative analysis of motor habits of eye movements in reading as related to length of the printed line, kind of material read, and level of reading ability. Tendency toward rhythmical motor habits was evaluated by means of a habit index, i.e., the percentage of successive lines having the same or nearly the same number of fixations. The habit index was higher for reading the shorter lines of print and slightly higher for the more efficient readers. This apparent relation is accounted for in terms of probability, however, rather than a habit for the fewer the fixations, the better the chance of two adjacent lines having the same number of fixations. The habit index favored neither the reading of easy narrative nor scientific prose. Another alleged sign of rhythmic eye movements or *short-lined motor habits* as they are sometimes called, is a characteristic temporal distribution of pause durations which is alleged to be fostered by shorter lines. Although some evidence for such a pattern was found, it was no different in short than in long lines of print. In view of his results. Sisson suggests that the concept of *short-lined motor habits* is useless. The reviewer would add that this notion of rhythmic reading is not only a useless concept but a harmful one. For years the view that rhythmical eye-movement behavior is characteristic of effective reading and is highly desirable has directed attention toward eye movements. This has led to an undue emphasis upon peripheral oculomotor mechanics to the sacrifice of adequate attention to the more important central factors of comprehension and assimilation (see *training of eye movements* p. 110). The findings of Walker (120) are in line with the above conclusions. He found "no evidence of a habit of moving the eyes any fixed distance or in any rhythm" in his analysis of eye movements of good readers. If rhythm is absent in the eye movements of highly skilled readers, as indicated in these studies, the view that rhythmical eye-movement patterns are desirable for effective reading becomes meaningless. Weaver (123) found a high degree of variability in ocular behavior characteristic of both memorizing and recall.

Motor efficiency. Tinker (111) has ascertained the relation of oculomotor efficiency to reading proficiency. There were no significant correlations between reading proficiency and either accuracy of visual fixation measured in terms of number and extent of corrective moves at fixation point or speed of convergence-divergence moves. Only when extremes of the group were compared was there a slight relationship evident. These findings refute previously held views. Similarly, Strombert (99) found no differences between fast and slow readers in convergence and divergence movements during reading. Knehr (51) noted that vergence movements in reading monocularly were the same as for binocular reading.

Eye-movement time. The proportion of time taken by interfixation movements in reading is relatively small and varies with kind of material read. Tinker (109) found that the per cent of reading time devoted to moves ranged from 5.3 to 9.6 with an average of 7.3. The more careful and analytical the reading, the smaller the relative time taken by moves. Luckiesh and Moss (63), after noting that there is no clear vision during horizontal eye movements in reading, state, "However, this unique and advantageous arrangement is stated to be absent in the case of vertical movements of the eyes." There is no sound basis for the latter statement.

Monocular and binocular vision. Apparently oculomotor behavior is about the same whether one or both eyes are used in reading. Knehr (51), employing eye-movement measures, found no significant differences in performance between monocular and binocular reading. Furthermore, the oculomotor patterns were alike under the two conditions. Clark (23) found that alternating vision which occurs during bar-reading, did not appreciably alter reading efficiency as measured by eye-movement records. Further experimenting with a larger group of subjects and with a more adequate sample of reading is needed to check these results. Nevertheless there is evidence to support the above findings. A mass of data indicates that central factors of comprehension and assimilation rather than peripheral, i.e., mechanical factors, are dominant in determining reading proficiency.

FACTORS IN VARIABILITY

Variation in subject matter. For some time it has been recognized that variation in subject matter produces changes in eye movements.

Additional information on the nature of the variation is contained in the following studies. Seibert (89) made an analytical study of eye movements of 8th grade pupils while reading mathematics, biography, adventure, physical science, history and geography. Differences in all eye-movement measures occurred in going from one subject matter to another, but these differences are not as significant as the author appears to believe. The critical ratios ranged from .00 to 2.53. It is admitted that some children maintained a definite pattern of eye movements throughout all materials. This indicates a lack of flexibility in reading habits. A similar analytical investigation was completed by Stone (98) with college freshmen as subjects. They read arithmetic, biology, English, educational psychology, physical science and social science. Although the critical ratios are not large, differences that tended toward significance were obtained for eye-movement measures in going from one kind of material to another. Implications of the findings were not discussed. Both of these studies may be criticised for faulty statistical interpretations.

Condition of reader. This category is concerned with such factors as maturity level, reader proficiency, stuttering and deafness. Buswell (14) has investigated the eye movements of adult readers. Measures employed for diagnosing reading maturity include:

1. The span of recognition or amount of print recognized in a single pause depends directly upon fixation frequency. A relatively wide span indicates greater maturity.
2. Rhythmic regularity of progress of eye movements along the line of print is considered symptomatic of a high degree of maturity in reading.
3. For any given kind of reading the more mature reader will make fewer regressive movements than the immature reader.
4. The more mature reader employs eye-movement pauses of briefer duration.

On all these measures the adult readers tended to be less mature than high school seniors. In fact the adults were only slightly better than 6th grade children in reading material within the comprehension range of both children and adults. Buswell emphasizes that reading habits should be flexible, i.e., they should adjust themselves to the kind of material read. Nevertheless, he failed to note that in general his adult readers revealed a marked lack of flexibility of oculomotor behavior in going from easy to more difficult material.

Anderson (2) compared the oculomotor patterns of 50 good and 50 poor readers to find how such readers adjust to changes in the difficulty of the reading material, and to reading for different purposes. All eye-movement measures distinguished good from poor readers. For both groups, fixation frequency, regression frequency and pause duration increased with the more difficult reading material. Good readers showed a greater flexibility in eye movements in adjusting to increasingly difficult material. Greatest irregularity in oculomotor patterns occurred in reading to remember details. Here the differences between groups was present but less pronounced. It was concluded that two important determinants of regularity in eye movements are difficulty of material read and purpose of reading. In a qualitative analysis of eye movements of good readers, Walker (120) discovered that on the average fixations were distributed evenly throughout the line for both easy and difficult material. Pause duration was greater in the first part of lines. About 80 per cent of lines were covered by fixations and the last fixation was nearer the end of the line than the first. Extent of forward shift of eyes was not related to duration of fixation at end of the shift.

There is not complete agreement concerning the oculomotor patterns of stutterers. While Moser (73) found that eye movements in silent reading are less effective for stutterers than for normal children, Hamilton (42) found no differences. The former considers that stuttering may cause reading disability while the latter states that stutterers do not need special treatment in the silent reading program. Since Hamilton's experiment was more adequately controlled, the more acceptable view is that eye movements in silent reading are the same for both stutterers and normal readers. In Moser's (73) study, silent fixation was the same for stutterers and normal speakers, but when stutterers talked during fixation, a variety of eye movements occurred. During oral reading, differences in oculomotor patterns were associated with stuttering spasms. These findings suggest that extrinsic eye muscles as well as other muscle groups are not under control during stuttering. Working with another kind of handicapped group, La Grone (56) found that the eye-movement patterns of deaf children differed from those of hearing children. This was especially true for developmental levels in reading. Although pause duration tended to decrease from grade to grade, fixation and regression frequency increased and then decreased. The most common kind of regression occurred at the beginning of lines, but there were fewer regressions for deaf than for hearing children. Since the groups studied were very small, these findings need checking.

Oral versus silent reading. A few studies have been concerned with eye movements in oral reading. Buswell (14) points out that schools have recognized that the processes of oral reading and of silent reading are psychologically different. Experimental work yields evidence on the degree of this difference. The eye movements of good, poor and unselected college subjects during silent reading were compared with eye movements during oral reading by Anderson and Swanson (**4**). The correlations for the eye-movement measures ranged from .50 to .70. Coefficients for the poor readers were somewhat higher. In silent reading there were fewer fixations and regressions and shorter pause durations. Going from poor to good readers, the differences in eye movements between oral and silent reading increased. Thus the eye-movement patterns tended to be more similar in the two kinds of reading for poor readers. These data show, therefore, that there are common elements in oral and silent reading, especially among poor readers. According to Buswell (**14**) silent reading of adults differs from oral reading in respect to the degree that vocalization is suppressed. Evidence from eye-movement measures while silently reading material designed to exaggerate vocalization supports this view. Under these conditions the eye movements of poor readers were like movements in oral reading; those of good readers were not.

In studying the eye movements of adults during oral reading, Buswell (**14**) found that the mean eye-voice span is an excellent index of the maturity of a person's oral reading habits. The greater the span the more mature the reading. An excellent study of eye movements of good and poor readers during oral reading is reported by Fairbanks (**31**). Both eye and voice records were made. As in the above study the eye-voice span reflected maturity of reading habits. Although difficulty of textual material determined variations in eye-movement measures, the good readers were less affected than the poor readers. No fixed pattern of eye movements was discovered. Analysis of regressions in relation to eye-voice span together with a comparison of the eye-movement patterns of good and poor readers led to the conclusion that faulty eye movements do not cause errors in reading but that the errors are central in origin.

Miscellaneous. Certain studies are not easily classified under the above groupings. Simpson (**92**) found no significant correlation between eye-movement measures and time spent in leisure reading. It is concluded without justification that eye movements are not improved unless specifically trained. In another study, Simpson (**91**) obtained correlations of −.38 to −.52 between fixation and regression frequency versus college achievement and mental ability. Pause duration showed no appreciable correlation with achievement and intelligence. Reading test scores correlated higher than eye-movement measures with achievement and intelligence. The inference that eye-movement training is desirable because the correlations are low is a misinterpretation of the data (see *training eye movements* p. 110). Imus, Rothney, and Bear (45) found that the correlations for eye-movement measures versus college aptitude, scholastic achievement and reading test scores were only .07 to .38. Their conclusion that eye-movement measures are unreliable and invalid is not justified (see *reliability* p. 94 and *validity* p. 95). In a similar study with children, Leavell and Sterling (58) found low to moderate correlations between eye-movement measures and intelligence. Taylor (105) obtained significant differences between eye-movement measures of children making normal school progress and of those failing. The grade norms given by Taylor are of doubtful value because an unreliable sample of eye movements was obtained. In fact Taylor was largely responsible for providing the inadequate samples of reading material employed by subsequent users of the Ophthalm-o-graph. They are too short to yield reliable records.

Peripheral vision. For years writers have been emphasizing the importance of peripheral vision in reading. Cues perceived in peripheral vision not only furnish premonitions of coming words and phrases, but they also guide the movements of the eyes to new fixations along the lines of print. La Grone (55) has studied the relationship between peripheral vision and eye movements in reading. Accuracy of perception in the left and in the right peripheral fields was obtained from tachistoscopic exposure. High perception scores in the left peripheral field tended to go with fast reading, few fixations and shorter pause duration. But high perception scores in the right peripheral field tended to go with slow reading, more fixations and longer pause duration. Regression frequency was not significantly related to perception scores. Higher perception scores either in right or left peripheral fields are said to be due to preference. Since vision in the right peripheral field is supposed to be important in reading, why do the more able readers have better perception to the left? The author misses the point that in tachistoscopically exposed material the long practiced habit of reading material from left to right is dominant. It is logical to infer that such a habit is more effective for good than for poor readers. Furthermore, a short exposure test of peripheral vision is not an adequate measure of peripheral vision as it operates in the normal reading situation where the successive fields of vision overlap. While the findings in this study are interesting, they do not provide a crucial test of the role played by peripheral vision in normal reading.

SPECIAL READING SITUATIONS

Spelling. In a follow-up of an earlier study on spelling (see **107**), Gilbert (**33**) obtained records for six children over a three-year period. The eye movements revealed growth characteristics in perceptual habits. Although there was little change in pause duration, fixation and regression frequency decreased and there was evidence of a more effective method of attack in studying the words. The more mature spellers differentiated between hard and easy sections of words and their ocular responses tended toward simple verification reactions. Gilbert and Gilbert (**35**) recorded eye movements before and after training for speed and accuracy of perception in learning to spell with 4th, 5th, and 6th grade pupils. Training resulted in less perception time per word and fewer fixations and regressions, but no change in pause duration. Thus the training resulted in a larger fixation span and less attention to details, but the general pattern of cross study and accuracy of spelling was maintained. Here is a good example of how training in perception habits with no attention to eye movements yields changed oculomotor patterns which reflect more effective central processes of apprehension and assimilation. In a later study, Gilbert and Gilbert (**36**) found that spelling is learned more effectively when words are studied in isolation than when encountered as critical words in reading. Eye movements revealed that when words were in text the reading was disrupted, especially for poor spellers. The authors' suggestion that a concentration of fixations may represent spelling reactions rather than comprehension or recognition difficulties would hold only if the readers were set to learn spelling as they were in this experiment due to a prior spelling test.

Reading music. The reading of music is a highly specialized oculomotor task which involves both vertical and horizontal eye movements. Lowery (**60**) notes that individual differences are in terms of pause duration rather than fixation frequency. Weaver (**124**) obtained photographic records of eye movements and of keyboard performance for 15 trained musicians during sight reading and playing of three selections at a piano. From one to two notes were read per fixation. A musical note is considered equal to a word in reading. Pause durations are longer in music reading than in word reading and vary with the complexity of note relations. The author's correlations of .82 to .89 between pause duration and reading time and of only .27 to .58 between fixation frequency and reading time are in contrast to word reading where fixation frequency yields the high correlations. This supports Lowery's (**60**) finding that pause durations provide the individual variation. According to Weaver (**124**), predominance of horizontal or of vertical progression of eye movements depends upon the nature of the music. Notes on the treble staff are usually but not always read ahead of notes on the bass staff. The eye-hand span was highly variable. Regressions appeared to be related to kind of music read. In another study, Van Nuys and Weaver (119) investigated eye movements of 12 experienced musicians while reading musical rhythms and melodies of 12-note phrases, each presented for a period of 2.8 seconds. Memory span equalled the number of notes correctly played for each phrase exposed. Regressions occurred more for the melodies than for rhythms while pause duration increased with increase in complexity of note relations for both melodies and rhythms. Pause duration tended to be longer for rhythms than for melodies and there were more notes correctly played per fixation for the rhythms. Findings in these studies of eye movements in reading music furnish material for a comparison with the reading of words:

1. Considering a note comparable to a word, the fixation span is similar in both kinds of reading.
2. Regressions serve somewhat the same purpose in both situations.
3. Increase in complexity of text brings increase in complexity of eye-movement patterns for both music and words.
4. Reading music involves both vertical and horizontal eye movements while reading words in text involves only the latter. There are of course minor vertical deviations of the eye in reading a line of print. And the return sweep to the next line involving both vertical and horizontal displacement, is common to both situations.
5. In reading words, fixation frequency is the best indicator of reading rate but in music pause duration is best.

HEREDITARY AND CERTAIN ENVIRONMENTAL FACTORS

Twin studies. Two studies of twin similarities in eye-movement patterns while reading prose have been reported. Morgan (72), comparing eye-movement measures for members of pairs in a well controlled experiment, found very low correlations (.04 to .24) between artificial pairs, higher (.24 to .53) for fraternal twins, and still higher (.61 to .72) for identical twins. These coefficients indicate more fundamental bases for individual differences in eye movements in reading than is furnished by a "habit" concept. The "underlying processes of assimilation" constitute a more logical explanation of the results. Hereditary factors play a significant role to the extent shown by the correlations. It is noteworthy that pause duration, which is least susceptible to training, showed the highest correlations for twins. The author correctly infers that, in treating poor reading, consideration should be given to capabilities of the individual rather than exclusively to certain ideal motor sequencies characteristic of oculomotor patterns of efficient readers. In a second study, Jones and Morgan (50) report additional evidence on twin similarities in eye-movement patterns. Median correlations for eye-movement measures in reading revealed a trend similar to that in the earlier study: artificial pairs, r = .105; fraternal twins, r = .435; and identical twins, r = .530. Three judges had considerable success in identifying on the basis of general appearance the eye-movement photographs of identical twin pairs when they were mixed with photographs from fraternal twins. Considering the many factors which may modify oculomotor patterns, it is significant that this identification was better than chance. These results, therefore, support the earlier conclusion that eye-movement patterns in reading are not determined wholly by training. Genetic factors are involved to an appreciable degree.

Emotion. Emotional disturbances appear to modify eye-movement patterns. Strongin, Bull, and Korchin (102) report that when readers expect a shock, visual efficiency in terms of eye movements decreased by about 10 per cent and binocular coordination became worse in 14 per cent of readers. Warren and Jones (122) obtained eye-movement records while subjects read in the laboratory and while they were exposed in a high place. No significant differences in fixation and regression frequency or pause duration were found. But unsteadiness of fixations appeared when in the high place, especially for acrophobic subjects.

Illumination. Luckiesh (61) and Luckiesh and Moss (64, 65) report the effect of variation in illumination on eye movements in reading. Records were obtained at the beginning and at the end of one hour of reading under one foot-candle and again under 100 foot-candles of light. Speed of reading was slowed up significantly greater under the one foot-candle of illumination in comparison with the 100 foot-candles. The difference was due largely to changes in pause duration. Although the authors condemn speed of reading as a measure of readability of print, they employ eye movements for this purpose, apparently unaware that eye-movement measures are speed of reading measures.

Readability. Eye movements have been used extensively by others to determine legibility or readability of print. Bell (8) found no difference in eye movements for reading typewriting versus manuscript writing, but a significant difference in favor of typewriting versus cursive, and in favor of manuscript versus cursive writing. Differences were mainly due to changes in fixation frequency rather than pause duration. Tinker and Paterson have completed eight eye-movement studies on the readability of type. In a long series of studios (78) these authors had compared speed of reading for various typographical variations. Significant differences were found in certain comparisons. The analytical eye-movement studies were undertaken to discover the specific patterns of eye movements related to the disclosed differences. The non-optimal text, which was read slower than the more readable material, was characterized by the following changes in eye-movement patterns:

1. All capitals (slower) versus lower case—more fixations (**113**).
2. Old English (slower) versus a modern type face—more fixations (**114**).
3. Ten point type: 19 versus nine pica (slower) line widths—more fixations and longer pause duration; 19 versus 43 pica (slower) line widths—more fixations and regressions plus longer pause duration (**79**).
4. Six point type: 13 versus five pica (slower) line widths—more fixations, fewer regressions and longer pause duration; 13 versus 36 pica (slower) line widths—more regressions and longer pause duration (**81**).
5. Type size: 10 versus six point (slower)—more fixations and longer pause duration; 10 versus 14 point (slower)—more fixations and longer pause duration (**80**).
6. Type sizes in optimal line widths: 11 versus eight point (slower)—more fixations and longer pause duration; 11 versus six point (slower)—longer pause duration (**82**).
7. Black print on white background versus red print on dark green background (slower)—more fixations and regressions plus longer pause duration (**115**).
8. Several variations: optimal versus non-optimal (slower) typography—more fixations and regressions plus longer pause duration (**83**).

These analytical studies of eye movements in relation to typographical variations indicate that non-optimal conditions may only increase fixations or pause duration, or both of these. Regressions may increase or decrease. Factors operating to change the oculomotor patterns include changes in visibility (small print, low contrast between print and background), in word form (large print, all capitals), prevention of optimal use of peripheral vision (very long and very short lines), or lack of familiarity with a type face (Old English).

Visual fatigue. Eye movements, along with other measures, have been employed in the search for adequate measures of visual fatigue. Dearborn (**26**) presents some tentative data which suggests that eye movements may reveal visual fatigue. After a long day of visual work consisting of proofreading and study, speed of reading was reduced and saccadic movements through relatively long angles were retarded. Since only two readers served as subjects, these trends need checking. According to Clark and Warren (**25**), during a 65-hour vigil there was no consistent trend in variation of fixations, regressions or pause duration for eye movements in reading. Although sporadic and temporary changes in binocular behavior occurred for certain individuals, there were no uniform trends as the period of wakefulness increased. Since compensation is possible, eye movements apparently failed to furnish an adequate picture of fatigue effects. Kurtz (54), however, claims that eye-movement records reveal fatigue after 30 minutes of severe visual exertion. It is difficult to evaluate his data as presented. The most thorough evaluation of eye movements in reading as measures of visual fatigue has been made by McFarland, Holway and Hurvich (69). Thirty minutes of either rapid saccadic exercise or 30 minutes of rapid convergence-divergence movements failed to produce any significant changes in eye movements while reading. Although their results were negative, the authors consider that eye movements ought to provide a sensitive indicator of visual fatigue if a more sensitive recorder were available. Such an apparatus would yield a more precise measure of fixation and of binocular coordination. But fixation and regression frequency and pause duration in reading would remain unchanged.

In general these studies reveal that eye movements in reading are insensitive indicators of visual fatigue. Only when the fatiguing task is extremely severe is there a suggestion that eye movements are modified. Two factors are probably preventing fatigue from affecting eye movements in reading: (a) All experimenters have used only 10 to 12 lines of text to be read before the camera. The visual mechanism can compensate enough to give normal performance for the short periods of time required to read such materials. (b) Furthermore, there is no adequate way to evaluate precisely the degree of comprehension in the short paragraphs of text ordinarily used. The reader, since there is a tendency to maintain one's oculomotor pattern, may read two paragraphs with similar eye movements but with different degrees of comprehension. An adequate check of eye movements as measures of fatigue can be obtained only with long samples of reading text equated for difficulty and so constructed that comprehension is satisfactorily measured or maintained constant.

Anoxemia. Definite and consistent modification of eye movements in reading occurs with anoxemia. McFarland, Knehr, and Berens (67, 68) found significant increases in reading time, i.e., perception time plus eye-movement time, and fixation frequency in oxygen concentrations corresponding to 15,000 and 18,000 feet. Of greater importance, perhaps, was the qualitative variation of ocular movements in oxygen want. That is, nystagmoid tendencies and general unsteadiness appeared and there was an accentuation of abnormalities such as muscle imbalance.

Furthermore, latent oculomotor defects frequently became apparent indicating that the procedure may have clinical significance. In general, altered oculomotor behavior yielded a sensitive measure of the early effects of anoxemia.

EYE MOVEMENTS IN THE READING CLINIC

Views concerning the usefulness of eye-movement records in the reading clinic vary. Kurtz (53) is probably over optimistic when he claims that the importance of a permanent objective record, i.e., eye movements, can not be over-estimated. The view of Clark (21) that eye-movement photography serves as a valuable objective auxiliary to other techniques is more moderate. Opposed to these is the view of Imus, Rothney and Bear (44, 45) who, in appraising eye movement photography as a clinical method, are convinced that the technique lacks both reliability and validity. (See criticism of this view under *reliability* p. 94 and *validity* p. 95.) Tinker (110), in an evaluation of eye-movement measures, notes purposes for which they may be used. He concludes, however, that eye-movement records are not indispensable in clinical diagnosis of reading status. An exaggerated stress is placed upon the study of eye movements in the clinical situation by Hamilton (41) and Sievers and Brown (90).

Orthoptic training. Peters (84) found that training with the Binocular Synchronizer and the Squint Korrector yielded a reduction in fixation and regression frequency. These changes, however, were similar to those produced by motivated reading practice alone. More extensive studies of this kind are reported by Parkins (75, 76, 77). Apparently there was marked improvement in reading with corresponding changes in eye movements due to orthoptic training. In this group of studies, the results are so presented that it is difficult to evaluate them. The reviewer is inclined to reserve judgment until check studies are made.

Visual abnormalities. Extensive clinical use of eye-movement records has been made by Clark. In an initial study (17), comparing the eye movements of individuals having normal binocular balance with those having muscular imbalance (exophoria), revealed no differences in either fixation frequency, regression frequency, pause duration, or time required to complete divergence movements at the beginning of the lines of print. For similar comparisons in later reports (18, 19, 22), Clark also found no difference in fixation and regression frequency but there were differences in vergence movements during reading. The exophoric readers made significantly greater divergence movements at the beginning of lines and required more time to complete these corrective divergence movements. It is suggested that the large divergence movements of subjects with muscular imbalance may cause enough fatigue to be important in remedial reading programs. Clark (**20**) reports similar trends for a case of high exophoria. After noting that his experiments reveal no striking differences in oculomotor patterns between normal readers and those with muscular imbalance, Clark (**24**) concludes correctly that such deficiencies are apt to cause undue fatigue and should be corrected. His belief, however, that visual deficiencies cause reading disability in spite of the negative evidence is not well taken. In a correlational study, Stromberg (**100**) found no relation between amount of vergence movements and phoria of eye muscles. This, however, does not negate Clark's contention that it is desirable to correct muscle imbalance for the individual case.

Remedial reading. In an extensive study of remedial reading at the college and adult levels, Buswell (**15**) employed the eye-movement camera to measure with precision the changes in perceptual habits which occurred during the experiment. He states, however, that an eye-movement camera should be considered a research rather than a clinical instrument. Simpler methods of measurement which are entirely satisfactory in an ordinary remedial program are available. Once a procedure in the reading clinic has been validated by eye-movement records, there is no need to recheck the procedure in other remedial groups.

TRAINING EYE MOVEMENTS

Problem. The early experimenters were concerned with identifying and describing oculomotor behavior during reading. As this field of investigation developed, differentiating factors in eye-movement patterns were more emphasized. That is, stress was placed upon developmental variation, differences between good and poor readers, variation with changes in subject matter read, etc. This eventually led to the practice of attempting to improve reading performance by training eye movements. The necessary assumption is that eye movements are important determinants of reading proficiency. Thus, if the retarded reader is trained to use eye-movement patterns similar to those which characterize efficient reading, his reading proficiency would presumably improve. This emphasis upon the mechanics of eye movements tended to direct attention to peripheral factors as determinants of reading performance rather than to the central or underlying processes or perception, apprehension and assimilation.

Methods. Techniques of training eye movements in reading vary. On the one hand are those, represented by Pressey (see Tinker, **108**) and Sievers and Brown (**90**), who try to train subjects to pace eye fixations at three regular intervals along a line of print. It is true, however, that Sievers and Brown suggest that eye-movement training is but one small part of a reading program. Nevertheless it is naïve to assume that if a line of print is separated into three groups of words, that paced eye movements will result in reading each group with one fixation. In fact, Dearborn and Anderson (**27**) have demonstrated that such an arrangement results in more than three fixations. Buswell (**15**) also notes that few subjects ever achieve three fixations per line. At the other extreme are those who attempt to increase the fixation span by employing techniques which approximate the normal reading situation. Thus, Dearborn and Anderson (**27**) have devised a flexible film projection method for teaching phrasing to increase the size of fixation span. Approximately the same principle is involved in the still-film projector method developed by Buswell (**15**) to increase span of recognition or fixation span. Successive phrases of text are exposed in proper sequence and spacing to lead the eye along in normal perceptual sequences. Timing and nature of textual material are flexible in both techniques. Between these extremes a variety of techniques have been employed to train eye movements. The most commonly used one is the metron-o-scope designed to develop "controlled reading." It is described by Taylor (**105**). This is a mechanically operated machine for exposing each line of print in three successive sections. Taylor (**105**) states that the fundamental idea of the metron-o-scope is to develop the mechanical, i.e. eye movement, and interpretative processes simultaneously. As noted above, such techniques do not reduce fixations to three per line.

Controlled reading. A large number of reports describe training of eye movements or controlled reading. Since Anderson (**1, 3**) Sisson (**95**), Kottmeyer (**52**), Nelson (**74**), Pilant (**85**) and Traxler (**117**) have evaluated or reviewed much of this material, detailed notice to the separate studies will not be given here. The experimental reports on controlled reading include studies by Broom (**10, 11**), Dearborn and Anderson (**27**), Dodson (**28**) Garver and Matthews (**32**), Hallock (**39**), Lauer, Henry and Fritz (**57**), Lee (**59**), Robinson (**86**), Simpson (**93**), Taylor (**105**), and Witzeman (**125**). Although some of these studies were done without adequate controls and lacked acceptable statistical treatment, they are rather uniform in showing considerable improvement in reading due to the training. Three studies in this area will be surveyed in more detail.

Sisson, (**96**), using matched groups of adult readers found that reading with intent to improve produced significant changes in eye movements comparable to the changes from eye-movement pacing. In a more elaborate study, Cason (**16**) employing matched groups of school children, found no significant differences in improvement in reading for the students trained in phrase reading or in eye movements by use of the metron-o-scope in comparison with those who did motivated library reading. That is, free library reading resulted in the same gains as programs stressing the oculomotor mechanics of reading. All made significant gains. Buswell (**15**) employed a film projector for controlled practice to increase the fixation span or span of recognition. This resulted in improved reading proficiency which was reflected in changed oculomotor patterns. Emphasis was upon apprehension of broader perceptual units so that the reader would be freed from the requirements of detailed visual perception. This permitted greater attention to meanings.

Evaluation. Eye-movement training may be evaluated as follows:

1. Many so-called procedures for training eye movements or for controlled reading result in improved reading efficiency.
2. The improved reading status is reflected in modified oculomotor patterns.
3. The improvement obtained by eye-movement training, with or without elaborate apparatus, is no greater than that resulting from motivated reading alone.
4. There is no adequate evidence that training eye movements as such improve reading (**107**). Experiments concerned with pacing eye movements and controlled reading usually involve other techniques and are never divorced from increased motivation. Buswell (**15**) flatly states "training eye movements does not increase reading ability."
5. The training of eye movements too often becomes a ritual and tends toward an overemphasis upon oculomotor mechanics to the sacrifice of adequate attention to the more important processes of perception, apprehension and assimilation. This training may result in a decrease in the flexibility and adaptability of reading habits which characterize good readers. According to Buswell (**15**), "The exploiting of machines and gadgets" to control reading "by persons who do not understand the psychology of reading seems at present to be adding greatly to this mechanistic folly." And Traxler (**117**) has "difficulty in seeing any justification for purely mechanical attempts to train pupils in better eye movement habits."
6. The reviewer agrees with Traxler (**117**) that controlled reading techniques, i.e., training eye movements, "need not be used in a purely mechanical way but that they might well be used to supplement, in a way that would have strong motivating force," a program in which reading comprehension is emphasized. The reviewer believes, however, that as long as gadgets are used by those with an inadequate understanding of the psychology of reading, we will continue to have the undesirable emphasis upon oculomotor mechanics.

CENTRAL VERSUS PERIPHERAL PROCESSES

Problem. Training of eye movements is intimately related to the problem of whether eye movements are causes or symptoms of reading proficiency. The view that eye-movements are causes of reading proficiency is based upon the discovery that when there is ineffective reading, there are many fixations and regressions and that good readers use few fixations and relatively few regressions. This trend developed until the point was reached where it was claimed that the good reader is good because he uses few fixations and rarely makes a regression. Thus, to develop a good reader presumably it is only necessary to develop effective oculomotor habits. This gradually led to a strong emphasis upon peripheral factors, especially among the large group of writers with an inadequate knowledge of the psychology of reading. A few writers, including this reviewer (**107, 108, 111**) began to question this emphasis. It was pointed out that the more important determinants of reading proficiency were the central processes of perception, apprehension and assimilation. These processes are reflected in oculomotor patterns because eye-movement habits are very flexible and appear to adjust themselves readily to any change in the perceptual processes involved in reading. It would appear that eye movements merely reflect efficient or poor reading performance rather than cause it.

Recent evidence and conclusions. There is really no evidence to support the view that eye movements determine reading proficiency. As shown in the above analysis, the results obtained from training eye movements do not support the view that peripheral factors are dominant although based upon the assumption that they are. On the other hand, there are numerous reports which indicate that the central processes are the important determinants in reading performance. Analysis in which the writers conclude that eye movements reflect the central processes of perception, apprehension and assimilation include: Anderson (**1, 3**), Traxler (**117**), and Sisson (**95, 97**). A variety of recent experimental studies of eye movements have led to the conclusion that oculomotor patterns reflect poor and efficient reading, not cause it. These include: Anderson (**2**), Fairbanks (**31**), Knehr (**51**), Rogers (**88**), Sisson (**94**), Swanson and Tiffin (**104**), and Walker (**120**), and Walker and Molish (**121**). For earlier studies see Tinker (**107, 108**). Changes in eye-movement patterns occur when difficulties of comprehension arise, with variation in reading purpose or attitude etc. It is now well established that oculomotor reactions are exceedingly flexible and quickly reflect any variation in the central processes of perception, apprehension and assimilation.

J-356
7.9.70

Evaluation. Careful consideration of the field would probably indicate that there need be no conflict in assigning proper roles to peripheral and central factors in reading. To read, one must perceive words in proper sequence and to do this the eye fixations must progress in appropriate patterns. In other words, oculomotor patterns are a means of achieving the perceptual sequences necessary for apprehension and assimilation of the textual material. The particular eye-movement pattern employed is conditioned by the nature of the central processes. Clear apprehension and rapid assimilation of the textual material is reflected in relatively few fixations and regressions. But ineffective perception and confused apprehension, accompanied by difficulties in assimilation, characteristically produce many fixations and regressions. The latter, however, should not be confused with the oculomotor pattern occurring in highly analytical reading where it is necessary to reexamine the text and to work out relationships. In such a case, as shown by Tinker, Frandsen and others (see Tinker, **107**), the complex eye-movement pattern may indicate effective reading rather than inefficient central processes. It seems possible, therefore, that the unfortunate dichotomy of central versus peripheral processes might be avoided to some degree by omitting use of the terms eye-movement patterns and rhythmical eye movements, and by employing instead the concept of perceptual sequences. The latter is more meaningful in discussing the reading process and it avoids the unfortunate mechanical implications of the former.

SUMMARY STATEMENT

Important progress in the study of eye movements in reading has been made during the past ten years. Many of the investigations have been of high quality. Nevertheless unfortunate trends have appeared. Many unqualified individuals, those without an adequate background in either experimental procedures or the psychology of reading have been doing "research" in eye movements. The results are deplorable. Furthermore, the exploitation of the eye-movement technique has accentuated the unfortunate emphasis upon peripheral determinants of reading proficiency and has led to wide use of eye-movement measurement in the clinical situation although the eye-movement camera should be considered a research rather than a clinical instrument. There are signs, however, that the conclusions from sound research and the views derived from adequate analysis of results in the field are beginning to make headway.

BIBLIOGRAPHY

1. Anderson, I. H. An evaluation of some recent research in the psychology of reading. *Harv. educ. Rev.* 1937, **7**, 330–339.
2. Anderson, I. H. Studies in the eye movements of good and poor readers. *Psychol. Monogr.*, 1937, **48**, 1–35.
3. Anderson, I. H. The ophthalm-o-graph and metron-o-scope evaluated in the light of recent research on the psychology of reading. *Teach. Coll. J.*, 1941, **12**, 60–63.
4. Anderson, I. H., & Swanson, D. E. Common factors in eye movements in silent and oral reading. *Psychol. Monogr.*, 1937, **48**, 61–69.
5. Arnold, D. C., & Tinker, M. A. The fixational pause of the eyes. *J. exp. Psychol.*, 1939, **25**, 271–280.
6. Auersperg, A. Blickbewegung und optischer Erfassungsakt. *Z. ges. Neurol. Psychiat.*, 1939, **165**, 209–213.
7. Bayle, E. The nature and causes of regressive movements in reading. *J. exp. Educ.*, 1942, **11**, 16–36.
8. Bell, H. M. The comparative legibility of typewriting, manuscript, and cursive script: II. Difficult prose and eye movement photography. *J. Psychol.*, 1939, **8**, 311–320.
9. Brandt, H. F. A bidimensional eye-movement camera. *Amer. J. Psychol.*, 1937, **49**, 666–669.
10. Broom, M. E. An experimental study of methods for improving reading mechanics in upper elementary grades. *El Paso Schools Standard*, 1940, **18**, 23–29.
11. Broom, M. E. Improvement of reading mechanics. *Texas Outlook*, 1940, **24**, 29
12. Broom, M. E. The reliability of the reading graph yielded by the ophthalm-o-graph. *Sch. & Soc.*, 1940, **52**, 205–208.
13. Broom, M. E., Duncan, A. A., Emig, D., & Stueber, J. *Effective reading instruction.* New York: McGraw-Hill Book Co., 1942.
14. Buswell, G. T. How adults read. *Suppl. Educ. Monogr.*, 1937, No. 45.
15. Buswell, G. T. Remedial reading at the college and adult levels. *Suppl. Educ. Monogr.*, 1939, No. 50.
16. Cason, E. B. Mechanical methods for increasing the speed of reading. *Teach. Coll. Contr. Educ.*, 1943, No. 878.
17. Clark, B. The effect of binocular imbalance on the behavior of eyes during reading. *J. educ. Psychol.*, 1935, **26**, 530–538.
18. Clark, B. A photographic study of the binocular behavior of the eyes having monocular and binocular abnormalities. *Trans. Amer. Acad. Optom.*, 1935, **9**, 168–178.
19. Clark, B. Additional data on binocular imbalance and reading. *J. educ. Psychol.*, 1936, **27**, 473–475.
20. Clark, B. A case of high esophoria studied by eye-movement photography. *Amer. J. Ophthal.*, 1936, **19**, 503–507.
21. Clark, B. Eye-movement photography as a diagnostic method in the determination of reading disability. *Amer. J. Optom.*, 1936, **13**, 121–129.
22. Clark, B. A photographic study of the binocular behavior of eyes having monocular and binocular abnormalities. *Amer. J. Optom.*, 1936 **13**, 441–451.
23. Clark, B. A photographic analysis of alternating vision during reading. *Amer. J. Ophthal.*, 1937, **20**, 1142–1144.

24. Clark, B. Binocular anomalies and reading ability. *Amer. J. Ophthal.*, 1940, 23, 885–891.
25. Clark, B., & Warren, N. A photographic study of reading during a sixty-five hour vigil. *J. educ. Psychol.*, 1940, **31**, 383–390.
26. Dearborn, W. F. On the possible relations of visual fatigue to reading disabilities. *Sch. & Soc.*, 1940, **52**, 532–536.
27. Dearborn, W. F., & Anderson, I. H. A new method for teaching phrasing and for increasing the size of reading fixations. *Pyschol. Rec.*, 1937, **1**, 459–475.
28. Dodson, W. L. Mechanical aids in a practical reading program. *Texas Outlook*, 1939, **23**, 13–14.
29. Ehlers, H. The movements of the eyes during reading. *Acta. Ophthal. Kbk.*, 1936, **14**, No. ½.
30. Ehlers, H. Om øjenbevaegelserne ved laesning. *Ugeskr. Laeg.* 1936, **98**, 1131–1133.
31. Fairbanks, G. The relation between eye movements and voice in the oral reading of good and poor silent readers. *Psychol. Monogr.*, 1937, **48**, 78–107.
32. Garver, F. M., & Matthews, R. D. An analysis of the results of speed drills with the metron-o-scope to increase reading rate. *J. educ. Psychol.*, 1939, **30**, 693–698.
33. Gilbert, L. C. A genetic study of growth in perceptual habits in spelling. *Elem. Sch. J.*, 1940, **40**, 346–357.
34. Gilbert, A. J., & Gilbert, D. W. Reading before the eye-movement camera versus reading away from it. *Elem. Sch. J.*, 1942, **42**, 443–447.
35. Gilbert, L. C., & Gilbert, D. W. Training for speed and accuracy of visual perception in learning to spell: a study of eye movements. *Univ. Calif. Publ. Educ.*, 1942, **7**, 351–426.
36. Gilbert, L. C., & Gilbert, D. W. The improvement of spelling through reading. *J. educ. Res.*, 1944, **37**, 458–463.
37. Gray, W. S. Summary of reading investigations. *J. educ. Res.*, 1936, **29**, 407–432; 1937, **30**, 553–576; 1938, **31**, 401–434; 1939, **32**, 481–517; 1940, **33**, 481–523; 1941, **34**, 401–443; 1942, **35**, 401–442; 1943, **36**, 401–444; 1944, **37**, 401–440; 1945, **38**, 401–429.
38. Gray, W. S. Reading. In W. S. Monroe (Ed.) *Encyclopedia of Educational Research.* New York: Macmillan Co., 1941. Pp. 891–926.
39. Hallock, E. Diagnostic and remedial work in a reading clinic. *Nation's Schools*, 1941, **28**, 28–29.
40. Halstead, W. C. A method for the quantitative recording of eye movements. *J. Psychol.*, 1938, **6**, 177–180.
41. Hamilton, J. A. *Toward proficient reading.* Claremont, Calif.: Saunders Press, 1939.
42. Hamilton, P. G. *The visual characteristics of stutterers during silent reading.* New York: Author, 35 Claremont Ave., 1940.
43. Hoffman, A. C., Wellman, B., & Carmichael, L. A quantitative comparison of the electrical and photographic techniques of eye-movement recording. *J. Exper. Psychol.*, 1939, **24**, 40–53.
44. Imus, H. A., Rothney, J. W. M., & Bear, R. M. *An evaluation of visual factors in reading.* Hanover, N. H.: Dartmouth Coll. Publ., 1938.
45. Imus, H. A., Rothney, J. W. M., & Bear, R. M. Photography of eye movements. *Amer. J. Optom.*, 1943, **20**, 231–247.
46. Jacobsen, O. I. An analytical study of eye-movements in reading vocal and instrumental music. *J. Musicol.*, 1941, **3**, 1–32.
47. Jacobsen, O. I. An analytical study

of eye-movements in reading vocal and instrumental music. *J. Musicol.*, 1941, **3**, 69–100.

48. JACOBSEN, O. I. An analytical study of eye-movements in reading vocal and instrumental music. *J. Musicol.*, 1941, **3**, 133–164.
49. JACOBSEN, O. I. An analytical study of eye-movements in reading vocal and instrumental music. *J. Musicol.*, 1942, **3**, 197–221; 223–226.
50. JONES, H. E., & MORGAN, D. H. Twin similarities in eye movement patterns. *J. Hered.*, 1942, **33**, 167–172.
51. KNEHR, C. A. The effects of monocular vision on measures of reading efficiency and perceptual span. *J. exp. Psychol.*, 1941, **29**, 133–154.
52. KOTTMEYER, W. On the theory of controlling eye-movements in reading. *Sch. & Soc.*, 1940, **52**, 496–498.
53. KURTZ, J. I. Photographing the eye movements during reading—an ophthalmographic technic. *Amer. J. Optom.*, 1936, **13**, 176–185.
54. KURTZ, J. I. An experimental study of ocular fatigue. *Amer. J. Optom.*, 1938, **15**, 86–117.
55. LA GRONE, C. W., JR. An experimental study of the relationship of peripheral perception to factors in reading. *J. exper. Educ.*, 1942, **11**, 37–49.
56. LA GRONE, T. G. An analytical study of the reading habits of deaf children. *J. exper. Educ.*, 1936, **5**, 40–57.
57. LAUER, A. R., HENRY, L. K., & FRITZ, J. F. Improving the reading ability of college students. *Psychol. Bull.*, 1938, **35**, 670.
58. LEAVELL, U. W., & STERLING, H. A. A comparison of basic factors in reading patterns with intelligence. *Peabody J. Educ.*, 1938, **16**, 149–155.
59. LEE, R. B. Value of metron-o-scope reading. *National Elem. Prin.*, 1939, **18**, 109–111.
60. LOWERY, H. On reading music. *Diopt. Rev. & Brit. J. Physiol., Optics.*, 1940, **1**, 78–88.
61. LUCKIESH, M. *Light, vision and seeing.* New York: D. Van Nortrand Co. 1944.
62. LUCKIESH, M., & MOSS, F. K. The extent of the perceptual span in reading. *J. Gen. Psychol.*, 1941, **25**, 267–272.
63. LUCKIESH, M., & MOSS, F. K. Reading a line of print. *Sch. & Soc.*, 1941, **53**, 612–615.
64. LUCKIESH, M., & MOSS, F. K. *Reading as a visual task.* New York: D. Van Nostrand Co., 1942.
65. LUCKIESH, M., & MOSS, F. K. The task of reading. *Elem. Sch. J.*, 1942, **42**, 510–514.
66. MCFADDEN, F. Eye movements. *Amer. J. Optom.*, 1938, **15**, 181–184
67. MCFARLAND, R. A., KNEHR, C. A., & BERENS, C. The effets of anoxemia on ocular movements while reading. *Amer. J. Ophthal.*, 1937, **20**, 1204–1219.
68. MCFARLAND, R. A., KNEHR, C. A, & BERENS, C. The effects of oxygen deprivation on eye movements in reading. *J. exp. Psychol.*, 1937, **21** 1–25.
69. MCFARLAND, R. A., HOLWAY, A. H., & HURVICH, L. M. *Studies of visual fatigue.* Boston (Soldiers Field): Graduate School of Business Administration, Harvard University, 1942.
70. MILES, W. R. An early eye-movement photograph. *Psychol. Monogr.*, 1936, **47**, No. 212, xxx–xxxvi.
71. MILES, W. R. The steady polarity potential of the human eye. *Proc. nat. Acad. Sci., Wash.*, 1939, **25**, 25–36.
72. MORGAN, D. H. Twin similarities in photographic measures of eye move-

ments while reading prose. *J. educ. Psychol.*, 1939, **30**, 572–586.

73. MOSER, H. M. A qualitative analysis of eye movements during stuttering. *J. Speech Disorders*, 1938, **3**, 131–139.

74. NELSON, E. A summary of investigations dealing with results obtained from teaching reading with the metron-o-scope. *El Paso Schools Standard*, 1943, **20**, 3–21.

75. PARKINS, G. A. *The Claremont experiments.* Amer. Res. Coun. Optom. Bull. No. 25. Ord, Nebr.: 1941.

76. PARKINS, G. A. *The diagnosis and elimination of visual handicaps preventing efficient reading.* Fall River Mass.: Distinguished Service Foundation of Optometry, 1941.

77. PARKINS, G. A. *Experiment II at Boulder.* Amer. Res. Coun. Optom. Bull. No. 28. Ord, Nebr.: 1942.

78. PATERSON, D. G., & TINKER, M. A. *How to make type readable.* New York: Harper and Brothers, 1940 (for sale by authors).

79. PATERSON, D. G., & TINKER, M. A. Influence of line width on eye movements. *J. exp. Psychol.*, 1940, **27**, 572–577.

80. PATERSON, D. G., & TINKER, M. A. Influence of size of type on eye movements. *J. appl. Psychol.*, 1942, **26**, 227–230.

81. PATERSON, D. G., & TINKER, M. A. Influence of line width on eye movements for six-point type. *J. educ. Psychol.*, 1942, **33**, 552–555.

82. PATERSON, D. G., & TINKER, M. A. Eye movements in reading type sizes in optimal line widths. *J. educ. Psychol.*, 1943, **34**, 547–551.

83. PATERSON, D. G., & TINKER, M. A. Eye movements in reading optimal and non-optimal typography. *J. exper. Psychol.*, 1944, **34**, 80–83.

84. PETERS, H. B. The influence of Orthoptic training on reading ability. *Amer. J. Optom. and Arch. Amer. Acad. Optom.*, 1942, **19**, 95–111; 152–176.

85. PILANT, R. Reading improvement programs. *Engl. J.*, 1942, **31**, 235–237.

86. ROBINSON, F. P. An aid for improving reading rate. *J. educ. Res.*, 1934, **27**, 453–455.

87. ROBINSON, F. P. The tachistoscope as a measure of reading perception. *Amer. J. Psychol.*, 1934, **46**, 132–135.

88. ROGERS, M. V. Comprehension in oral and silent reading. *J. Gen. Psychol.*, 1937, **17**, 394–397.

89. SEIBERT, E. W. Reading reactions for varied types of subject matter: An analytical study of eye movements of eighth grade pupils. *J. exp. Educ.*, 1943, **12**, 37–44.

90. SIEVERS, C. H., & BROWN, B. D. *Improving your eye movements in reading (with methods manual).* Wichita, Kan.: McGuin Publ. Co., 1940.

91. SIMPSON, R. G. The relationship of certain functions to eye-movement habits. *J. educ. Psychol.*, 1942, **33**, 373–378.

92. SIMPSON, R. G. Does the amount of free reading influence the student's control of his eye movements in reading ordinary printed matter? *J. educ. Psychol.*, 1943, **34**, 313–315.

93. SIMPSON, R. H. Further analysis of the results of speed drills with the metron-o-scope. *J. educ. Psychol.*, 1941, **32**, 386–388.

94. SISSON, E. D. Habits of eye-movement in reading. *J. educ. Psychol.*, 1937, **28**, 437–450.

95. SISSON, E. D. The role of habit in eye movements in reading. *Psychol. Rec.*, 1937, **1**, 159–168.

96. SISSON, E. D. Eye-movement training as a means of improving reading ability. *J. educ. Res.*, 1938, **32**, 35–41.

97. SISSON, E. D. The causes of slow reading: an analysis. *J. educ. Psychol.*, 1939, **30**, 206–214.

98. STONE, L. G. Reading reactions for varied types of subject matter: an analytical study of the eye-movements of college freshmen. *J. exp. Educ.*,1941, **10**, 64–77.

99. STROMBERG, E. L. Binocular movements of the eyes in reading. *J. Gen. Psychol.*, 1938, **18**, 349–355.

100. STROMBERG, E. L. The relation of lateral muscle balance to the convergence and divergence movements of the eyes during reading. *J. Gen. Psychol.*, 1938, **19**, 437–439.

101. STROMBERG, E. L. The reliability of monocular photography in the investigation of reading. *J. educ. Psychol.*, 1942, **33**, 118–127.

102. STRONGIN, E. L., BULL, N., & KORCHIN, B. Visual efficiency during experimentally induced emotional states. *J. Psychol.*, 1941, **12**, 3–6.

103. STROUD, J. B. A critical note on reading. *Psychol. Bull.*, 1942, **39**, 173–178.

104. SWANSON, D. C., & TIFFIN, J. Betts' physiological approach to the analysis of reading diabilities as applied to the college level. *J. educ. Res.*, 1936, **29**, 433–448.

105. TAYLOR, E. A. *Controlled reading.* Chicago: Chicago Univ. Press, 1937.

106. TIFFIN, J., & FAIRBANKS, G. An eye-voice camera for clinical and research studies. *Psychol. Monogr.*, 1937, **48**, 70–77.

107. TINKER, M. A. Eye movement, perception, and legibility in reading. *Psychol. Bull.*, 1936, **33**, 275–290.

108. TINKER, M. A. Eye movements in reading. *J. Educ. Res.*, 1936, **30**, 241–277.

109. TINKER, M. A. Time taken by eye-movements in reading. *J. genet. Psychol.*, 1936, **48**, 468–471.

110. TINKER, M. A. An evaluation of eye-movement measures for optometrists. *Amer. J. Optom.*, 1938, **15**, 1–5.

111. TINKER, M. A. Motor efficiency of the eye as a factor in reading. *J. educ. Psychol.*, 1938, **29**, 167–174.

112. TINKER, M. A. Reliability and validity of eye-movement measures of reading. *J. exp. Psychol.*, 1939, **19**, 732–746.

113. TINKER, M. A., & PATERSON, D. G. Influence of type form on eye movements. *J. exp. Psychol.*, 1939, **25**, 528–531.

114. TINKER, M. A., & PATERSON, D. G. Eye movements in reading a modern type face and old English. *Amer. J. Psychol.*, 1941, **54**, 113–115.

115. TINKER, M. A., & PATERSON, D. G. Eye movements in reading black print on white background and red print on dark green background. *Amer. J. Psychol.*, 1944, **57**, 93–94.

116. TRAXLER, A. E., & SEDER, M. *Ten years of research in reading.* New York: Educ. Records Bureau, 1941.

117. TRAXLER, A. E. Value of controlled reading: summary of opinion and research. *J. exper. Educ.*, 1943, **11** 280–292.

118. VALENTINE, W. L., TROYER, M. E., & BROWN, M. A. *Eye movements in reading.* (Film) Columbus: Ohio State Univ., 1936.

119. VAN NUYS, K., & WEAVER, H. E. Memory span and visual pauses in reading rhythms and melodies. *Psychol. Monogr.*, 1943, **55**, No. 1, 33–50.

120. WALKER, R. Y. A qualitative study of the eye movements of good readers. *Amer. J. Psychol.*, 1938, **51**, 472–481.

121. WALKER, R. Y., & MOLISH, H. The relationship between factors of

ocular efficiency and eye-movement measures at the college level. *J. educ. Psychol.*, 1941, **32**, 145–150.

122. WARREN, A. B., & JONES, V. Effect of acrophobia upon reading ability as measured by reading comprehension and eye movements in reading. *J. genet. Psychol.*, 1943, **63**, 3–14.

123. WEAVER, H. E. Eye movements in memorizing and recall. *Psychol. Bull.*, 1937, **34**, 730–731.

124. WEAVER, H. E. A survey of visual processes in reading differently constructed musical selections. *Psychol. Monogr.*, 1943, **55**, No. 1, 1–30.

125. WITZEMAN, B. E. An experimental study using the ophthalm-o-graph and metron-o-scope in the diagnosis and treatment of reading defects. *J. Psychol.*, 1941, **11**, 307–334.

126. YAMAMOTO, S. An experiment on the eye-movements in the reading of the Japanese language. *Jap. J. Psychol.*, 1935, **10**, 773–789.

COLOR ADAPTATION TO 1945*

JOZEF COHEN
Cornell University

It is the purpose of this paper to present a critical history of the phenomenon known to workers in color as color adaptation, color fatigue, or color minuthesis.

Although usually attributed to later investigations, the first scientific observation of color adaptation appears to be that described by the French mathematician and physicist De La Hire (**12**, p. 289) in 1694. His was a description of observations of the sun, the course of perceptual change, and the resulting after image. He ventured an hypothesis concerning the fatigue of visual receptors, and although it is somewhat erroneous in view of later investigations, it is characteristically the Hering theory.

The first important investigation was the 1863 contribution of Maria Bokowa (**5**). Bokowa wore copper oxide glasses of 1 mm. thickness from four to five hours. The glasses had a transmission which was predominantly red, but they could not have been highly selective since she was able to see colors in all parts of the spectrum. The experiment was controlled, in a manner, since a spectrum was viewed from time to time during the course of adaptation and comparisons made from one observation to another. Further, there was an attempt at quantification by viewing a color wheel and describing the perceptual color which corresponded to a physical stimulus resulting from an arbitrary number of circular degrees of arbitrary "primaries." In general, Bokowa's results may be summarized as follows:

1. After color adaptation had taken place, all colors appeared as either yellow or blue.
2. The brightest color was a mixture of yellow and blue.
3. The cut off point of the visible spectrum shifted toward the short end of the spectrum.

It is to be noted that nowhere does Bokowa indicate that the adaptation phenomena continued to a neutral gray although Helson and Judd (**19**, p. 383) refer to Bokowa as having reached complete adaptation. Bokowa (**5**, p. 163) refers to red as appearing less intensive during adaptation and this indicates that some change in intensity may have taken place. She may, however, have been referring to a loss in saturation although the word saturation does not appear in the paper. Bokowa, then, reports changes in hue, and makes references to changes in intensity.

---

* From a doctoral dissertation *The Color Adaptation of the Human Eye* directed by Professor H. P. Weld. For an additional paper based on the same dissertation, see Cohen (**11**).

In 1865, the phenomena was noted by Aubert (**3**) in his observations of red and blue colored squares against achromatic backgrounds. The loss of saturation was referred to as (**3**, p. 131) "Adaptation der Netzhaut für Farben." He adds that adaptation for blue light is faster than adaptation for red light. It is interesting to note that Sheppard (**33**, p. 35) says, "Nowhere in his writings does Aubert connect this term [adaptation] with color-changes, for which he uses the term 'fatigue'." The use of the term adaptation for both white and colored light apparently did originate with Aubert.

Sigmund Exner's experiment (**14**) in 1868 was the first which used a spectrum as adapting and reacting stimuli. His experimental procedure was to throw a spectrum on a screen. Slits were placed in the screen so that the light of the adapting color came through one slit while the light of the reacting color came through the other. Exner was, of course, studying the effect of the after image as well. By observing at a distance the beam of light which penetrated the first slit, he adapted a small area of the retina. The adaptation time was ten seconds. The eye was then brought very close to the second slit which had the effect of stimulating a much larger area of the retina. And in this larger area there appeared the after image of the light stimulus of the first slit. Inspection of Exner's table shows the following results:

1. When the stimulus color and the adapting color are identical, the adapted portion of the retina is always lower in saturation than the non-adapted portion.
2. The adapted portion of the retina perceived all colors as red, green, or blue.

With respect to this second result, Exner attempted to locate the spectral position of these basic colors whose hue does not shift during adaptation and toward which all other hues shift. For this purpose he devised an instrument utilizing a telescope and a spectrum. His results are as follows:

1. For red the position is from the long end of the visible spectrum to a point between C (656 mμ) and D (589 mμ), perhaps somewhat closer to C.
2. For green the location is between E (527 mμ) and b (517 mμ).
3. For blue it is in the immediate environment of the line G (431 mμ).

One of the few early studies directly concerned with the changes in intensity during color adaptation was performed by Schön (**31**) in 1874. Schön was not interested in adaptation phenomena as such but rather in the physiological mechanism to which discrepancies among previous experimenters who had determined the differential sensitivity of the retina to the different portions of the spectrum could be attributed. Schön suggested that the differences between the many investigators might be due to the color adaptation of the retina.

Schön used prisms to obtain spectral stimuli. These stimuli were presented to an observer through a tube which was divided in half, forming a stimulus and comparison field. The intensity of each half was controlled by a variable slit, the width of the slit being a measure of the intensity. The experiment was begun by making the slits equal on both halves of the tube. The comparison field was shielded from the view of the monocular observer while the stimulus was presented for 3, 5, 10, or 15 seconds. At the end of the interval the comparison field was presented to the subject who stated whether they were equal in intensity. If they were not judged to be equal, the slit of the comparison field was changed and the determinations continued until they were judged to be equal. Three stimulus colors were used—red, green, and blue. Schön came to these conclusions:

1. After five seconds exposure to spectral red, green or blue, one half of the brightness is lost.
2. There is a further decrease in brightness which takes place gradually.
3. This further decrease is greatest for blue and least for red.

Many criticisms have been leveled at this experimental procedure and consequently the results. His two test fields were not strictly comparable because of the construction of the apparatus. He ignored the differential sensitivity of the retina to various wave lengths in comparisons of heterogeneous wave lengths. For other criticisms and an alternative procedure which avoids most of the defects, see Almack (**2**, p. 27 ff).

Ssamujlow's 1889 experiment is only reported in abstract form (**34**). So meager is the information available that it is difficult to determine the exact nature of his investigation. Ssamujlow considered the duration of the after image as a measure of the adaptation of the retina, a theory which may not be strictly true. Ssamujlow does not adequately explain what he means by adaptation, but presumably he refers to loss in saturation. Using spectral light sources, he arrived at the following results:

1. The retina does not adapt equal amounts for different lights. It is most for red, less for green, and still less for blue.
2. The degree of the adaptation is dependent on the time and the strength of the stimulus.
3. The peripheral retina adapts to a lesser degree than the fovea.

There are two contributions to the literature by Carl Hess. The first (20), published in 1890, describes his experimental procedure. A telescope was arranged in conjunction with spectroscopes so that there appeared in the telescope a left and right field, either of which could be illuminated with monochromatic light. The field on the left was a square while the field on the right had the form of a circle. The telescope could be shifted so that either the right or left field came into view. The observer's eye was placed at the telescope and fixated the right edge of the left square field. After a given interval the telescope was shifted to the right field with the cross hairs in the center of the circle. Thus only one half of the adapted retina is covered with the new fixation, so that changes which occurred by virtue of the adaptation process are apparent by comparison of the fields. Nine different homogeneous lights and two non-spectral lights made by mixing spectral red and spectral violet were used as both adaptation and reaction stimuli. Changes in hue were noted for all combinations of lights taken two at a time; each light was used as an adaptation light and matched with all other lights as reaction stimuli. There were two adaptation times—10 seconds and 35 seconds.

The experiment was continued in 1893 (21) and here the apparatus was modified to give results in terms of wave length. The spectroscope was fitted with two adjustable slits; through one slit passed the adapting light while the reacting light passed through the other. The other spectroscope was fitted with a single adjustable slit. The telescope was not kept in a fixed position. The left half of the telescopic field could be filled with either the reacting light or the stimulus light while the right field was filled with light from the adjustable single slit of the right spectroscope. The observer fixated a point between the two fields for 50, 70, 75, or 180 seconds. The left field was illuminated with the stimulus light while the right field showed blackness. At end of the adaptation time the stimulus light was shut off and the reacting light and the comparison light presented simultaneously. The comparison light was readjusted after successive trials until it matched the perception of the left field. Hess reaches these conclusions:

1. There is a tendency for colors to appear as either blue, green, or red to the color adapted eye. This is only a tendency and the data present an erratic appearance. Hess' cursory attempt to locate the wave lengths of these colors using only three colors as stimuli (having theoretical implications in the Helmholtz theory) found them to be 490 mμ for blue, 565 mμ for green, and 650 mμ for red. These figures are quoted erroneously by Almack (2, p. 26) and Sheppard (34, p. 35 ff). To be sure, the basic colors are not well identified, but the data do indicate a red basic.

2. The shift in hue increases with time. There is no reference to saturation in this paper.

Voeste's 1898 experiment (**45**) is one of the keystones in the history of color adaptation. His apparatus was the Helmholtz color mixer (**18**, p. 395 ff) as described by König and Dieterici (**22**). Essentially two fields were presented to the subject. The observer monocularly fixated a point between the left stimulus field and the right comparison field. The stimulus field was illuminated with a monochromatic source while the comparison field was unilluminated. After ten seconds, the observer matched the right field with what was seen on the left. It was matched for hue by variation of the prismatic surfaces and for intensity by variation of the slit openings, while saturation was completely disregarded. The stimulus fields covered the visible spectrum at three or four different levels of intensity. Voeste's results may be summarized as follows:

1. All colors tend to shift in hue toward three basic hues, 560 mμ, 494–498 mμ, and 460–470 mμ. Subsequent experiments using these wave lengths as test patches show no shifts in hue. Troland suggests that these changes are due only to the Bezold-Brücke effect (**42**, p. 210 or **43**, p. 180).
2. The comparison field is from 18.9% to 83.3% of the intensity of the stimulus field. These figures are in terms of slit openings and, of course, fail to take account of the differential sensitivity of the retina.

In 1899 and 1900 Burch wrote two papers on the production of artificial color blindness. The results of the experiments described in these papers are so unusual that extreme caution needs be exercised in interpretation. In the first paper (**8**, p. 2) Burch wrote, "[The method] consists of exposing the eye to bright sunlight in the focus of a burning glass behind a transparent screen of the proper colour, and keeping it there until all sensation of that particular colour is lost." This must mean, it appears, that adaptation goes to a neutral gray. Burch continues (**8**, p. 3), "With a combination of ruby glass . . . it takes from a few seconds to two or three minutes to produce complete red blindness." It is not clear as to whether he means the perception was of red or of neutral gray. He does claim that for an eye stimulated as described above, the sensation of red is not present, and the eye is completely blind to that hue. If this is the case, adaptation not only reaches a neutral gray, but the eye remains in that state for some time since its perception of red would be similar to its perception of gray. This condition, Burch continues, does not in any way affect the perception of other colors having no red component. It has no effect on intensity at all.

In his second paper (**9**) Burch attempted to show that the effect was not pathological since it could be produced by moonlight. He wrote (**9**, p. 213), "In producing red-blindness, as there are no colors to the left of red, I have generally used a part of the spectrum which excites only the red sensation, and which therefore must continue to appear visible at all. But I have not thought it desirable to push the fatigue of the retina far enough to destroy this sensation of light." Of course, the spectrum could appear as gray and still be visible.

Beck (**4**) observed the effects of intense illumination on colors and noticed the adaptation effect. He continued his experiments with spectral sources and reported that the colors were not visible as hues or were desaturated, depending on the degree of adaptation.

Porter and Edridge-Green (**29**) adapted the eye with a band of color for twenty seconds. Immediately after the adaptation period the eye was shifted to view a spectrum larger than the band; the band filled only the center portion of the spectrum so that comparison could be made between the adapted and unadapted portions. Although the experimental procedure is interesting, their results are erratic and their experiments cursory. They noted that the retina did not lose sensitivity for red, but sensitivity was lost for blue. In general their results follow no general form or law and there is no direct statement pertaining to the law of color adaptation.

In 1913, however, Edridge-Green (**13**) contributed an additional series of experiments of more significance, two of which we report:

I. After the eyes have been exposed to sodium light for twenty minutes, a spectrum is examined. The yellow has completely disappeared from the spectrum; the green and red meet without any intermediate color. But the red, orange, and green have not lost any of the yellowness which they previously possessed.

II. Blue-green spectacles were worn for ten minutes. At first all white objects appeared as a vivid blue-green, but at the end of ten minutes' time, a white object appears white. On examining the spectrum with these glasses before adaptation, there was no red to be seen, there were small amounts of orange, and the blue, green, yellow, and violet were visible. After adaptation, there were no marked changes in the orange or any other color, except the green which took on a paler and more yellow appearance. The region of the spectrum corresponding to the dominant region of the transmission of the glasses was most affected by adaptation, while the blue and yellow on each side appear bluer and yellower respectively.

The conclusions follow:

1. No color is seen by the color adapted eye unless the corresponding physical stimuli are present in the light reaching the retina. Color adaptation pro-

duces its effect by subtraction, and not by the addition of any new color sensation.

2. Color adaptation increases with time.

3. The dominant wave length corresponding to the adaptation stimulus color of spectral regions appears colorless. Colors immediately on either side of the dominant wave length are shifted higher and lower respectively.

The experimental apparatus used by Brückner (7) was Helmholtz' *Spiegelapparat* (18, p. 161). This was essentially a piece of plateglass placed perpendicular to the table. The observer is able to look through the glass to the surface of the table beyond it, or to see the reflection of the near surface of the table, or both, depending on how the glass is covered. Brückner placed an adapting stimulus composed of two parts —one half chromatic and the other half achromatic—at the far end of the table. All of the colors were from the Hering series. The chromatic colors covered the visible spectrum and are indicated in Ångstrom units. The chromatic half was either white, gray, or black. Each hue appeared three times—once with each of the achromatic colors. The reacting stimulus, placed on the near table, was one of the seven hues used as one half of the adapting stimulus or one of six additional colors. The experiments were carried out binocularly using daylight. The individual fixated the adapting stimulus through the plate glass. After ten to fifteen seconds the apparatus was adjusted so that he observed the reacting stimulus as a reflection. The results are expressed in terms of whether or not the reacting stimulus is more effected by the previous adaptation to the chromatic or achromatic stimuli. Since Brückner did not report for the case when the reaction color was the same as the stimulus color, no conclusion may be drawn except that changes in saturation do occur.

In 1920 Hubert Sheppard (33) made the first systematic investigation of color adaptation. His paper presents four main experiments:

I. Five observers who fixated in direct sunlight seven large sheets of Hering colored paper which covered the range of the visible spectrum, described the course of adaptation during fixation. In general, they report either no hue change or a hue change to yellow, with no change in intensity. The loss in saturation proceeded to a neutral gray in from 45.9 seconds to 220 seconds. When the subjects observed 180° of the same colors mixed with 180° of black and white on a color wheel, complete adaptation took place in from 23.4 seconds to 121.3 seconds.

II. In the second experiment, circles of the Hering papers placed in a field of Hering gray were illuminated by a 75 watt bulb. By means of a telegraph key attached to a kymograph the observers signaled when important perceptual changes occurred and also when uncontrolled eye movements occurred. They did not give a running account of the adaptation process. In general, adaptation was not characterized by shifts in hue as in the first experiment nor was complete adaptation reported by all subjects. All observers, however, reported loss in saturation and told of a cloud-like fog which moved over the color. The adaptation times (times at which no further change occurred) ranged from 40.3 seconds to 120.2 seconds. Sheppard determined the chromatic limens for each of the stimulus hues by the method of limits using a disc containing black, white, and the hue. The limen has a negative correlation with the adaptation time. Wherever the adaptation time is long, the chromatic limen is small.

III. In this series a remodeled spectrophotometer so designed that the subject, observing through an artificial pupil, was able to perceive a small portion of the spectrum as a disc, was used. Five positions were chosen from the visible spectrum. All but a single observer reported complete adaptation. The time ranged from 76.6 seconds to 169 seconds. Since Sheppard noticed that these times formed a curve which closely resembled the brightness curves for spectral colors, he performed another series in which the colors were equated for brightness value. The resulting figures, although more similar than those obtained before the brightness values were equated, still showed another factor to be operating. This, Sheppard reasoned, was the chroma value already demonstrated as being related to the adaptation time in his second experiment.

An additional series was conducted in which the chroma values were made equal (by subjective techniques). It does not appear that the brightness values were kept constant as well. His apparatus was too crude to control both factors at the same time. The adaptation times (for complete adaptation) for the various colors were now close enough so that Sheppard considered chroma as being the chief determiner of adaptation times. Introspective reports indicated no hue change except for red and violet which shifted toward yellow and blue respectively. There was no reported change in intensity.

IV. The last experiment was an attempt to determine whether complete adaptation times could be reached with very high intensities. A carbon arc provided the light source while gelatin filters were placed over a pair of glasses. The experiment was monocular. The subject fixated the arc while wearing the glasses until no further change took place. He then fixated a spectroscope which was equipped with a powerful illuminating source and a rheostat to control the intensity. (The rheostat no doubt changed the color temperature.) This served as a comparison light. The subjects reported complete adaptation in from 105.3 seconds to 197 seconds. There was some shift in hue toward yellow but most subjects reported no shift in hue at all. As Almack points out (2, p. 20 ff), Sheppard's description of this portion of his experiment is not clear. It is difficult to determine the nature of this comparison light formed by the spectroscope.

Troland (41) has reported the results of what he terms photopic adaptation. These are the same experiments referred to in his previous papers (37, 39, 40). The eleven experiments will be described:

I. In the first experiment, designed to test the equilibrium sensation for direct sunlight on white paper, a large white paper was fixated in direct sunlight for ten minutes. He concluded that prolonged exposure of the retina to large, very intense, achromatic stimuli does not involve a reduction of the sensation to neutral gray. The luminosity of the perception after adaptation is much greater than that of mid-gray. Further, the equilibrium sensation just described is brought about rapidly.

II. In this experiment large sheets of colored paper, selected to cover the range of the visible spectrum, were observed in direct sunlight. From these experiments he drew the conclusions that after exposure of the retina to large chromatic stimuli of high intensity, the resulting perception is not one in which the hue of the stimulus has completely disappeared. There may be some shift in hue but the perception is never one of neutral gray. The equilibrium point is reached soon after fixation.

III. A dark gray square, placed in the center of a yellowish-green Hering paper, was fixated until the green component of the paper faded leaving a light blue. At the end of four minutes a shutter arrangement substituted a Hering violet paper for the gray square. The outlying field became a darker blue while there was no reappearance of the green. But on the introduction of a white object into the green adapted area, the green color returned. Additional tests indicate that this phenomenon is not due to the contraction of the pupil. The experiment, of course, lends support to the theory of central control of color adaptation.

IV. Observations similar to those of Experiment I were made with stimuli consisting of two juxtaposed Hering papers of different hues. One paper filled the upper half of the field while the other filled the lower half. The six possible combinations of the four *Urfarben* were employed. Troland concluded that the principle of simultaneous color contrast applies to the equilibrium position. The introduction of a neutral object into the bi-chromatic field brings about the same change in quality for both halves of the field which was described in the previous experiment for a simpler stimulus. An experiment using a bipartite fixation of black and white failed to show any tendency of the field to approach neutral gray.

V. This experiment concerns the perception of a small luminous spot on a dark background. If the illumination is low, the spot tends to disappear and remain below the visibility threshold or fluctuate between visibility and invisibility. As Marbe (25) demonstrated, the more intense the spot, the less is the tendency for it to disappear. This was confirmed by Troland, and with spots of colored light fixated monocularly at high intensities, observers reported that the stimulus colors changed as in the previous experiments, but it did not approach a neutral gray nor did the spot fluctuate.

VI. The fluctuations of the last experiment are explained by supporters of the Hering theory as being due to eye movements permitting the retina to recover. See, for example, Ferree (**15**). A large sheet of Hering paper was fixated. The paper was large enough to be well out of the field of possible eye movements. With steady monocular fixations, repeated fluctuations in the field could be observed. Temporary disappearance could be obtained.

Fluctuations described here cannot be explained on the basis of adaptation followed by recovery due to eye movement. Deviations in fixation would move the image on the retina through not much more than one fiftieth of its diameter. Changes in accommodation could not be responsible for fluctuations since such changes would only be on the edge of the image.

VII. This experiment was designed to investigate whether the changes in the iris can be responsible for the visual fluctuations which the previous experiment showed were not due to eye movements or accommodation. The stimulus surface described in the previous experiment was perforated and a telescope placed behind. The iris of the observer was illuminated so that changes could be observed through the telescope. The observer pressed a telegraph key when the colored spot was seen to disappear. The pupillary contractions and expansions have a perfect correspondence with the reported fluctuations.

VIII. If changes of the iris are responsible for fluctuative disappearances, it is obvious that the elimination of these changes by the use of drugs such as pilocarpin or atropin or the use of an artificial pupil should cause the elimination of these fluctuations. Drugs have been used on similar problems by Pace (**28**) and McDougall (**27**). Both investigators report that fluctuations were present. For the effects of the artificial pupil, see Troland (**38**). The intensity of the colored spot stimuli was placed under the control of the subject. Starting with a very low intensity, he increased the brightness until no fluctuations occurred for three minutes. The stimulus was then regarded as being at the required threshold. The results indicate that if the action of the iris is eliminated, no fluctuations occur for long intervals. The fluctuations which do occur, according to Troland, are due to mechanical difficulties of the artificial pupil, to the electrical condition of the visual system, binocular rivalry, etc.

IX. Spots of various colors equated for brightness were fixated with the artificial pupil. Since the brightnesses were of moderate intensity, no fluctuations of the spot stimuli were observed. The adaptation did not proceed to a neutral gray. If the stimulus is a psychological primary, there is no essential shift in hue. Intermediate hues tend to lose the red and green components.

X. The subject observed a colored spot for twenty seconds. The intensity of the spot was then decreased until the subject indicated that it had disappeared. Fixation was kept constant by means of the bright crescents appearing at the edges of the adapted area. Reappearance of the whole area always occurred.

XI. In the greater number of visual studies it is necessary to stimulate but a single eye while the other is blindfolded. This is an ideal condition for binocular rivalry. To demonstrate this phenomenon, one eye observed a colored spot. Immediately after fixation, the stimulus was transferred to the other eye where it appeared at first in its proper color. But soon it became a sparkling white, reminding one of binocular luster. It is therefore suggested that the after image appeared in the other eye and when mixed with its complement (the stimulus hue), it appeared white.

Troland did not discover this phenomenon. It was first noted by Newton in a letter to Locke and is the subject of Titchener's thesis. See Titchener (**35**) or for a brief discussion in English (**36** p. 49 ff). See also Franz (**16**, p. 44) for evidence that this effect does not occur.

Troland and Langford reported in abstract form some experiments concerning chromatic minuthesis (**44**). Minuthesis is their term for adaptation. They contend that changes which occur are only changes in saturation. Any further changes in hue are due to the Bezold-Brücke effect.

In 1923 Hamilton and Laurens (**17**) used a comparison technique to study adaptation. One eye was fatigued for thirty seconds. Then both eyes were placed at telescopes which presented a colored field to each eye. The fields were adjusted until a just noticeable hue discrimination was indicated. The results were as follows:

1. Red, green, and blue have a selective effect when used as stimulating sources. All other colors do not.
2. During adaptation brightness is reduced.

One of the more important studies of chromatic adaptation was performed by Almack (**2**) in 1928. Her problem was to measure the loss of sensitivity of the color adapted eye. Spectrum lights, used as stimuli, were controlled with respect to hue, saturation, and intensity. The threshold of the spectral light which was used to induce loss of sensitivity was determined at the end of each adaptation period. This was accomplished by the introduction of a motor rotated sectored disc into the path of the stimulus light. The rotating disc could be removed at will. Determinations were made after varying amounts of time (from 2 to 300 seconds) and gathered together in the form of curves. Comparison could be made between different intensities of the same hue, but not between different hues since the thresholds for the different wave lengths have different values. Comparison is made, however, when the determinations are reduced to a percentage basis. The experiments were made with both a dark adapted and light adapted eye. Allmack drew the following conclusions from her work:

1. For stimulus lights made equal in intensity, the sensitivity of the non-adapted eye is from greatest to least, yellow, green, blue, red. Initial loss in sensitivity is greater than at any other period. Initial sensitivity is greater for the dark-adapted eye than for the light-adapted eye, but both sensitivities tend to become equal with time. The rate of adaptation increases with the intensity of the stimulus.
2. For lights equalized for perceived saturation of the stimulus, the rates of adaptation to the various stimuli are very unequal. The largest rate (2, p. 115) "of adaptation by no means occurs at that intensity of light which gives maximum saturation to the color sensation. . . . The rate of adaptation is a function of the physical intensity of light, not of its subjective aspects, saturation and brightness."
3. For lights made equal with respect to wave length, the loss of sensitivity varies with the wave length. Under dark-adaptation, chromatic sensitivity is greater than for light-adaptation.

There is some confusion as to just what is measured by the technique described in this experiment. Consider a retina adapting to a stimulus light of a particular hue, saturation, and intensity. After a time, we determine the intensity limen of the eye to the same hue and saturation. This is true since we use the same source and introduce a rotating sectored disc. If we consider the color solid to be a double cone, then in dealing with hues of maximal saturation which lie on the surface, the saturation of such hues is dependent on intensity. They are dependent variables. Then if the intensity of our stimulus light is reduced, we have also reduced the saturation. On the other hand, if the color solid is considered to be a cylinder (and this is implicit in the 'fundamental sensation curves' of König and Dieterici (22)) there is no relation between saturation and intensity. Under these conditions, measurements by the above sectored disc can only indicate changes in perceived intensity.

Roaf (30), too, used the threshold of the eye as a measure of sensitivity. His apparatus was very complicated, but essentially it permitted the threshold to be determined after stimulation. He concluded that when any light stimulates the retina, the sensitivity, even of another part, is decreased to the short wave lengths and somewhat less to the long.

Kravkov (23) utilized the Helmholtz color mixer (18, p. 395 ff) in

1928 to test a theory of Lasareff concerning adaptation. The left field was fixated for ninety seconds and then matched with the right. The fixation was monocular. The match was made for intensity only. The results, claims Kravkov, support the outlined theory and indicate curves described by an equation with variable parameters. Kravkov's curves are, however, misleading. For example the first curve on page 93 (23) would fit the observation points just as well if it were completely inverted, or, indeed, if just a horizontal line were drawn. The conclusions to which Kravkov came are the following:

1. The decrease of intensity of a color sensation following color adaptation is of great importance and accords a lessening of the stimulus to 50% of its initial value.
2. The course of adaptation depends not only on the intensity of the stimulus but also on the hue. At the same intensity, violet (451 mμ) is more quickly adapted than red (656 mμ) and less than green (550 mμ).

Helson and Judd (**19**) performed two important experiments in 1932. The first concerned itself with eye movements and their relation to color adaptation. An apparatus was constructed so that an individual placed his head in a sphere 36 inches in diameter. The inside of the sphere was lined with orange-red Hering paper. The subject fixated the inside of the sphere binocularly. Strong illumination eliminated pupillary changes. In theory the apparatus was so constructed that any eye movements did not change the effective stimulus. Observations were made for from five to seventy-five minutes. The conclusions arrived at by Helson and Judd by use of their 'adaptation sphere' are as follows:

1. The neutral condition of gray did not appear permanently. Gray was perceived but its perception was cyclic and spasmodic. They think that this was probably due to the pupillary effect.
2. The red component vanished making the perception more yellow.
3. Most of the adaptation took place within the first few minutes of stimulation.

In the second experiment, the subjects wore strongly selective red or green glasses (the transmissions are given in detail). The observer wore these glasses for as long as five hours and at hourly intervals identified colored Hering papers for hue, saturation, and intensity. It is difficult to understand how a subject wearing glasses transmitting nothing less than 600 mμ could see the blue component in a blue paper. Note this report (**19**, Table 2, p. 388). The results are as follows:

1. After five hours, the lighter colors appeared in the hue of the glassse while the darker colors were either reutral gray or complimentary in hue to that of the glasses (because of the after image).

2. At first there was a sharp loss in saturation as well as a shift in hue—a shift toward yellow.

Attention should be called to the theoretical explanations in this paper. Helson and Judd assume that if complete adaptation did occur while the individual is wearing colored glasses, he would see everything in 'natural' colors. This, they point out, is impossible if the glasses are highly selective since adaptation cannot supply what is physically not present. Further, they continue, if the glasses are not highly selective, the objects may appear in 'natural' colors since the predominant wave length will cause quick adaptation to itself and since the glasses transmit at least some of all the wave lengths, objects will be seen in 'natural' colors. This last is, of course, quite possible. Such non-selective glasses might lead to such a result. But adaptation with highly selective glasses can occur, at least theoretically. In this case everything transmitted would eventually be perceived as a neutral gray; perceived as a gray physical stimulus would be perceived by a non-adapted eye.

In 1934 Bouma (6) commented that when looking at a white surface with a black card at the center with sodium (or neon) light, the black part, after a time, took on a violet (or green) hue. His study, however, was more concerned with after images.

The first approach to the color adaptation problem by use of twin colorimeters was by Wright (47) in 1934. The apparatus was that described by him in (46). It can be characterized by the following. There were two eye pieces each of which was connected to a Wright colorimeter. Any hue could be seen in either of these eye pieces by adjusting three knobs associated with each colorimeter. The knobs mixed three spectral primaries. All hues in high intensities but not all saturations could be produced. The third eye piece, to the right of the other two, presented a field of white light. The subject was dark adapted for a period of thirty minutes. He then fixated the fields of the two colorimeters; the left field was seen by the left eye while the right field was seen by the right eye. Binocular rivalry did not occur, according to Wright. The test patch of the right eye was matched with the comparison field of the left. The calibration readings were noted by the experimenter. The subject then shifted his right eye to the eye piece showing white light and fixated this patch for three minutes. The left eye, during this time, was unilluminated. At the end of three minutes, the observer again fixated the colorimeter fields as before. The fields no longer match. The observer quickly matches the left patch with the perception of the right. When the match is completed, he signals the experimenter who notes the time and the wedge readings, and the subject proceeds to make another match. The process is continued until no further change is noted.

This attack of the problem is somewhat different from anything else we have encountered. Here, it is the recovery time after adaptation that is measured. And recovery is given in terms of the spectral primaries of the colorimeter. At the beginning the two fields of the colorimeter are the same. The right eye is stimulated with white light which, it is assumed, depresses (or adapts) the activity of the receptor for the test patch to a level different from the original perception. The eye is then allowed to recover to the level of the test patch. This recovery is supposedly measured with the left eye. What is not taken into account, however, is that the eye does not recover to the level set by the test patch but continues to be stimulated by it until the receptors reach the equilibrium position. It appears that the white light is a confusing element and we wonder that Wright did not adopt the experimental procedure used in Cohen's (**11**) experiment. For additional papers using retinal recovery as a measure of adaptation see Allen (**1**) and the bibliography included in that paper.

Since the results of Wright's experiment are given in specialized response curves, no interpretation can be given until they are mathematically transformed to a more widely used system of primaries.

Color fatigue of the peripheral retina was investigated by Cogan and Cogan (**10**) in 1938–39. A perimeter was employed to hold the colors. The test color (Hering paper) was framed in a background having the same saturation as the test color. The time between the first perception of the color and its eventual merging into the background was considered to be a measure of adaptation. The results indicated the following:

1. Color adaptation continues to a neutral gray. The rate of adaptation increases toward the peripheral field.
2. Green was the most easily adapted color while red was the most difficult.
3. Dark adaptation influences color adaptation times.

There is a cursory study by Schouten and Ornstein (**32**) published in 1939. Their experiments utilize a separate field for each eye. They report that it was difficult to keep the two fields from fusing or overlapping. The comparison could be made in brightness only. Their contribution is that the adaptation of a test patch (here loss in intensity) is dependent upon the state of the rest of the retina.

Cohen (**11**) has recently completed an investigation in which he utilized an instrument which presented two test patches to a binocular subject. The apparatus was so constructed that any hue could be placed in each patch by the adjustment of three controls. A test patch was placed in the right side of the instrument, a neutral gray on the left. After one minute the observer matched the left with what he saw on the right. The match being made, the left side is returned to gray while the right continues to adapt. After another minute, another match is obtained. In this way successive matches are made against time for a particular test patch. The instrument readings which are given in terms of monochromatic radiations of red, green, and blue (700 mμ, 546 mμ, 436 mμ) are transformed to the monochromatic system of hue, saturation, and intensity. Thus *quantitative* changes in these latter variables were plotted against time. He did not employ intermediate colors.

Under the conditions of Cohen's experiment the following three laws may be stated:

Law I. The course of color adaptation consists of a gradual but never complete loss of saturation, with no change in hue, and an increase in intensity.

Law II. The rate and degree of desaturation depends upon the hue, the relative intensity, and the relative saturation of the stimulus color.

a. The loss of saturation for a green stimulus is most rapid, it is less for red, and least for a blue stimulus.

b. The less intense the stimulus, the greater the loss of saturation.

c. The less saturated the stimulus, the more rapid is the initial rate of desaturation; the equilibrium position, or point of lowest desaturation, is independent of the saturation of the stimulus.

Law III. The rise in intensity is dependent on the saturation and intensity of the stimulus and independent of the hue of the stimulus.

a. If the saturation and intensity of the stimulus are both high or both low, the rise in intensity will be greater than if one is high and the other is low.

SUMMARY OF METHODS

The methods used by the various investigators may be summarized as follows:

I. Introspective—mere description of adaptation.

De La Hire (12), Aubert (3), Ssamajlow (34), Sheppard (33), Troland (41), Helson and Judd (19), Bouma (6)

II. *Controlled introspective—perceived objects are carefully controlled (no comparison with unadapted portion of retina).*

Bokowa (5), Burch (8, 9), Beck (4), Edridge-Green (13), Sheppard (33), Helson and Judd (19)

III. *Liminal–limen (intensity) is measured after adaptation.*

Almack (2), Roaf (30)

IV. *Comparative monocular introspective—two fields viewed monocularly—differences are noted introspectively.*

Exner (14), Porter and Edridge-Green (29)

V. *Comparative monocular non-introspective—two fields viewed monocularly—match made on instrument.*

Exner (14), Schön (31), Hess (20, 21), Voeste (45), Edridge-Green (13), Kravkov (23)

VI. *Comparative binocular introspective—two fields viewed binocularly—differences noted introspectively.*

Brückner (7)

VII. *Comparative binocular non-introspective—two fields viewed binocularly, one field for each eye—match made on instrument.*

Wright (47), Hamilton and Laurens (17), Schouten and Ornstein (32)

VIII. *Comparative duo-binocular non-introspective—two fields viewed binocularly, both fields seen by both eyes—match made on instrument.*

Cohen (11)

SUMMARY OF RESULTS

The results obtained by the various investigators may be summarized as follows:

I. *Saturation*
Partial loss: Bokowa? (5), Aubert (3), Exner (14), Troland (41), Helson and Judd (19), Almack? (2), Roaf? (30), Cohen (11)
Complete: Edridge-Green (13), Sheppard (33), Cogan and Cogan (10)

II. *Intensity*
No change: Burch (8, 9), Sheppard (33)
Loss: Schön (31), Voeste (45), Almack (2) Hamilton and Laurens (17), Roaf (30) Kravkov (23), Schouten and Ornstein (32)
Increase: Cohen (11)

III. *Hue Shift*

Investigator	R	Y	G	B	V
			*Basic Colors**		
Bokowa (**5**)		x		x	
Exner (**14**)	656–589		527–517		431
Hess (**20, 21**)	650	565		490	
Voeste (**45**)		560		498–494	470–460
Sheppard (**33**)		x		x	
Troland (**41**)	x	x	x	x	
Helson and Judd (**19**)		x			
Cohen (**11**)	no shift				

BIBLIOGRAPHY

1. Allen, F. On reflex visual sensations and color contrast. *J. opt. Soc. Amer.*, 1923, **7**, 913–942. Also in *Amer. J. physiol. Opt.*, 1925, **6**, 339–374.
2. Almack, M. R. A quantitative study of chromatic adaptation. *Psychol. Monogr.* No. 174, 1928, **38** (2), 1–118.
3. Aubert, H. *Physiologie der Netzhaut.* Breslau: E. Morgenstern für Aug. Schultz und Co., 1865.
4. Beck, A. Ueber künstlich hervorgerufene Farbenblindheit. *Pflüg. arch. ges. Physiol.*, 1899, **76**, 634–640.
5. Bokowa, M. Ein Verfahren, künstliche Farbenblindheit hervorzubringen. *Zeitschrift für Rationelle Medicin*, 1863, (Reihe III) **17**, 161–165.
6. Bouma, P. J. Phenomena of eye fatigue and after images with monochromatic light. *Physica*, 1934, **1**, 429–436.
7. Brückner, A. Über die Sättigungsänderungen von Pigmentfarben durch Ermüdung der Netzhaut mit farbig wirkendem Lichte. *Arch. Augenheilk*, 1919, 85, 12–27.
8. Burch, G. J. On artificial temporary colour-blindness, with an examination of the colour sensations of 109 persons. *Philos. Trans.*, 1899, **191** (B), 1–34. Also in *Proc. roy. Soc.*, 1898, **63**, 35–38.
9. Burch, G. J. On the production of artificial colour-blindness by moonlight. *Proc. roy. Soc.*, 1899–1900, **66**, 216–219.
10. Cogan, D. G., & Cogan, F. C. Color fatigue in the peripheral visual field. *Ophthalmologica*, 1938–1939, **96**, 137–154.
11. Cohen, J. Color adaptation in man. *Amer. J. Psychol.*, 1946, **59**, January (in press).
12. De La Hire, M. *Memoires de Mathematique et de Physique.* Paris: L'Imprimerie Royale, 1694.
13. Edridge-Green, F. W. Colour adaptation. *Proc. roy. Soc.*, 1913, **86** (B), 110–115.

---

* Basic colors indicate those hues which do not change and toward which all other hues shift. An x indicates an investigator concluded that a particular basic color is present. In some cases the notation is in millimicrons.

14. Exner, S. Ueber einige neue subjective Gesichtscheinungen. *Pflüg. arch. ges. Physiol.*, 1868, **1**, 375–394.

15. Ferree, C. E. An experimental examination of the phenomena usually attributed to the fluctuation of attention. *Amer. J. Physiol.*, 1906, **17**, 81–120.

16. Franz, S. I. On after images. *Psychol. Rev. Momogr. Sup.* No. 12, 1899, **3** (2), 1–61.

17. Hamilton, W. F., & Laurens, H. The sensibility of the fatigued eye to differences in wave-length in relation to color blindness. *Amer. J. Physiol.*, 1923, **65**, 569–584.

18. Helmholtz, H. v. *Treatise on Physiological Optics.* (3rd ed.). (Vol. II). (Ed. J. P. C. Southall). Menasha: George Banta Publishing Co. for Optical Society of America, 1924.

19. Helson, H., & Judd, D. B. A study in photopic adaptation. *J. exp. Psychol.*, 1932, **15**, 380–398.

20. Hess, C. Ueber die Tonänderungen der Spectralfarben durch Ermüdung der Netzhaut mit homogenem Lichte. *v. Graefes Arch. Ophthal.*, 1890, **36** (1), 1–32.

21. Hess, C. Ueber die Unvereinbarkeit gewisser Ermüdungserscheinungen des Sehorgans mit der Dreifasertheorie. *v. Graefes Arch. Ophthal.* 1893, **39** (2) 45–70.

22. König, A., & Dieterici, C. Die Grundemfindungen in normalen und anomalen Farbensystemen und ihre Intensitätsvertheilung im Spectrum. *Z. Psychol.*, 1892, **4**, 241–317. Also in *Gesammelte Abhandlungen zur Physiologischen Optik von Arthur König.* Leipzig: Johann Ambrosius Barth, 1903. Pp. 214–321.

23. Kravkov, S. W. Über die Adaptation des Auges an farbige Lichtreize. *J. Psychol. Neurol., Lpz.*, 1928, **36**, 87–102.

24. Landauer, J. *Spectrum Analysis.* (2nd ed.). (Trans. J. B. Pringle). New York: John Wiley and Sons, 1907.

25. Marbe, K. Die Schwankungen der Gesichtsempfindungen. *Philos. Stud. (Wundt)*, 1893, **8**, 615–637.

26. McDougall, W. Some new observations in support of Thomas Young's theory of light- and colour-vision. *Mind*, 1901, (New Series) **10**, 52–97; 210–245; 347–382.

27. McDougall, W. The physiological factors of the attention process. *Mind*, 1902, (New Series) **11**, 316–351 and *Mind*, 1903, (New Series) **12**, 289–302; 473–488.

28. Pace, E. Zur der Schwankungen der Aufmerksamkeit nach Versuchen mit der Masson'schen Schiebe. *Philos. Stud. (Wundt)*, 1893, **8**, 388–402.

29. Porter, A. W., & Edridge-Green F. W. Negative after-images and successive contrast with pure spectral colours. *Proc. roy. Soc.*, 1912, **85** (B), 434–439.

30. Roaf, H. E. The influence of coloured lights on the sensitivity of the eye to various regions of the spectrum: a study in relation to theories of colour vision. *Quart. J. exp. Physiol.*, 1928, **18**, 243–262.

31. Schön, W. Einfluss der Ermüdung auf die Farbenempfindung. *v. Graefes Arch. Ophthal.*, 1874, **20** (2), 273–286.

32. Schouten, J. F., & Ornstein, L. S. Measurements on direct and indirect adaptation by means of a binocular method. *J. opt. Soc. Amer.*, 1939, **29**, 168–182.

33. Sheppard, H. Foveal adaptation to color. *Amer. J. Psychol.*, 1920, **31**, 34–58.

34. Ssamujlow (no initials). Zur Frage nach Ermüdung der Netzhaut durch verschiedene Farben (Roth, Grün, und Blau). (In III Congress russischer Aertze zu St. Peters-

burg). *St. Petersburger Medicinische Wochenschrift*, 1889, **14** (3), 26–27.

35. TITCHENER, E. B. Ueber binoculare Wirkunge monocularer Reize. *Philos. Stud.* (*Wundt*), 1892, **8**, 231–310.

36. TITCHENER, E. B. *Experimental Psychology*, (Vol. I, Part II). New York: Macmillan Co., 1924.

37. TROLAND, L. T. Adaptation and the chemical theory of sensory response. *Amer. J. Psychol.*, 1914, **25**, 500–527.

38. TROLAND, L. T. The theory and practice of the artificial pupil. *Psychol. Rev.*, 1915, **22**, 167–176.

39. TROLAND, L. T. Apparent brightness: its conditions and properties. *Trans. Illumn Engng Soc.*, 1916, **11**, 947–966.

40. TROLAND, L. T. Preliminary note: the influence of changes in illumination on after images. *Amer. J. Psychol.*, 1917, **28**, 497–503.

41. TROLAND, L. T. The colors produced by equilibrium photopic adaptation. *J. exp. Psychol.*, 1921, **4**, 344–390.

42. TROLAND, L. T. Vision I. Visual phenomena and their stimulus correlations. In C. Murchison (Ed.), *The Foundations of Experimental Psychology*. Worcester: Clark Univ. Press, 1929. Pp. 169–215.

43. TROLAND, L. T. *The Principles of Psychophysiology*. (Vol. II). New York: D. van Nostrand Co., 1930.

44. TROLAND, L. T., & LANGFORD, C. H. The stability of hue under chromatic minuthesis. *J. opt. Soc. Amer.*, 1925, **10**, 294–295.

45. VOESTE, H. Messende Versuche uber Qualitätsänderungen der Spectralfarben in Folge von Ermüdung der Netzhaut. *Z. Psychol.*, 1898, **18**, 257–267.

46. WRIGHT, W. D. A trichromatic colorimeter with spectral primaries. *Trans. Opt. Soc., Lond.*, 1927–1928, **29**, 225–242.

47. WRIGHT, W. D. The measurement and analysis of colour adaptation phemomena. *Proc. roy. Soc.*, 1934, **115** (B), 49–87.

THE EFFECTS OF NOISE*

F. K. BERRIEN
Colgate University

The development of modern industrial society with its concentration of population and machinery in confined areas has brought with it a concomitant concentration of noise. The popular science literature of the past twenty years is dotted with articles describing the deleterious effects of noise in highly emotional terms unwarranted by the facts. Few problems have attracted a wider range of professional interests than the problem of noise and what to do about it. Physicians, public health authorities, architects, psychologists, otologists, physicists, sound and electrical engineers have contributed to the literature. On the industrial side interest and concern is evidenced by the telephone and radio industries, automotive, airplane and railroad equipment designers, manufacturers of office equipment, building materials and a wide range of consumer appliances. Most recently the safety engineers and insurance companies have expressed an interest because of the growing recognition of occupational deafness. One is reminded also of the various noise abatement campaigns which sprang into prominence in the early 1930's.

In the light of this wide spread interest, one might reasonably expect to find a considerable body of well grounded scientific literature relevant to the topic of noise and how it affects the behavior of man. On the contrary there is a paucity of studies dealing with the problem at the human level in terms which provide a basis for generalizations even though a wealth of unsystematized experience by acoustical engineers points to the conclusion that people are benefited by noise reduction. Part of the reason for this state of affairs lies in the fact that only within the past 20 years has there been any attempt to standardize terminology and techniques in the measurement of noise itself. Even today the available procedures are still in the developmental stage where many improvements are to be desired.

It is the purpose of this review to summarize critically those studies pertaining to the effects of noise on human beings. Occasionally it may be necessary to refer to studies of animals to point up certain areas but in general the emphasis will be upon studies using human subjects.

---

* This review has been prepared as a preliminary step for a research program dealing with the effects of noise reduction on manufacturing personnel, sponsored by the Acoustical Materials Association. The author is especially indebted to Hale J. Sabin and Clarence W. Young for assistance in the preparation of this paper.

Consideration will be given in turn to the effects of noise on output speed and accuracy, fatigue and energy expenditure, the course of noise-adjustment, stimulation deafness, and finally the factors determining relative annoyance. Preliminary to an examination of these problems it is necessary to summarize the methods of measuring noise and their limitations.

DEFINITION OF, AND PROCEDURES IN MEASURING NOISE

Noise has been defined as unwanted sound (60). This statement bearing as it does a heavy value judgment appears to be widely accepted among physicists, engineers and the like. Paradoxically the psychologists lean toward a definition couched in more physical terms. Thus Dockeray (11) writes, "If a complex of vibrations is not in a harmonic ratio, we speak of the result as noise." And Dashiell (7) agrees in saying, "When air vibrations are non-periodic and irregular, or are less than two full vibrations, they produce *noises* rather than *tones*." The present paper is primarily concerned with noise as unwanted sound recognizing at the same time that the degree of annoyance may vary from a theoretical zero to some maximum point depending upon factors to be enumerated later.

The unit of noise measurement most widely used is the decibel. This is a logarithmic unit based on sound energy measurements. Two sounds differ by N decibels when $N = 10 \log_{10} E_1/E_2$. (E_1 = sound energy density of sound 1; E_2 = sound energy density of sound 2.) In practice it has been found that 1 decibel represents *approximately* 1 J.N.D. in a 1000 cycle tone. The zero point on the decibel scale could be set at any value but has now been standardized at 10^{-16} watts per square centimeter which is a convenient value somewhat lower than the normal threshold of hearing for a 1000 cycle tone. Publications prior to the tentative adoption of standard accoustical terminology in 1936 were often confusing in their use of the decibel scale because the minimal auditory threshold varies with frequency of vibration. Early writers refer to a tone of 200 cycles at 30 db without specifying whether they mean 30 db above the 1000 cycle tone threshold or 30 db above the 200 cycle threshold.

Less frequently used in this country is the word *phon* which is numerically equivalent to a decibel when considering the standard 1000 cycle tone. For any tone other than a 1000 cycles its phon value is determined by equating in loudness that tone with the standard. Referring to Figure 1 this means that a 300 cycle tone having an intensity level of 40 db. above 10^{-16} watts per square centimeter will appear to be equal in loudness to a 1000 cycle tone having a 30 db. intensity level above the same base. Hence the 300 cycle tone at that intensity has a loudness level of 30 phons.

Related to this unit are the terms *loudness units* and *sones*. If a 1000 cycle tone at 40 db. be taken as a standard and a second tone be reduced to one-half the subjective loudness of the standard, the second is said to be one-half a sone. Conversely a tone adjusted so that the normal observer judges it to be twice as loud as the standard is said to be 2 sones.

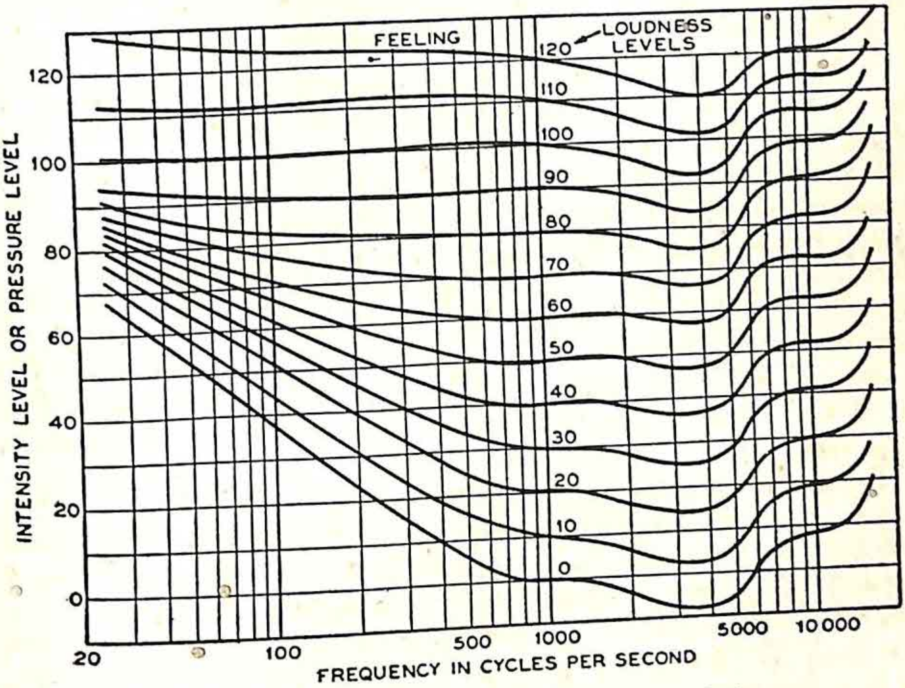


FIGURE 1. EQUAL LOUDNESS CONTOURS (30)

Stevens proposed this latter unit and has published curves (51) showing the relation of sones to db. level for a number of representative frequencies. The sone curve for tones between 700 and 4000 cycles is almost identical with a curve of loudness units (L.U.) published by Fletcher (16) and subsequently adopted with slight modification by the American Standards Association (60). The L.U. curve was constructed on the basis of several independent studies conducted in this country and in England where estimates of various fractional and multiple loudnesses were plotted against phons. The chief difference between the L.U. scale and the sone scale is in the magnitude of the numerical values assigned, the L.U.'s being roughly 1000 times greater than sones. Either scale is more indicative of the subjective effects of noise level changes than the db., phon or the direct intensity scales. Although the American Standards Association has specifically defined decibels, phons, and loudness units, it is common practice at least in speaking and often in writing to use the term decibel to cover physical sound intensity, equal loudness contours, or subjective loudness. It would be more meaningful to persons unfamiliar with acoustical terminology if phons were translated into either sones or loudness units.

One practical way by which proper usage of these terms could be encouraged is by redrawing the equal loudness contours of Figure 1 (a figure that appears in almost every acoustical discussion) in terms of either sones or L.U.'s. As the figure stands the spaces between the contours do not represent equal increments in loudness. Stevens and Davis (51) have already made a step in this direction by assigning equivalent sone values to the contours as already drawn. The translation into L.U.'s can be made by using Table I.

TABLE I

EQUIVALENT PHON AND L.U. VALUES

Phon	*L.U.*	*Phon*	*L.U.*
110	215000	50	2200
100	88000	40	980
90	38000	30	360
80	17100	20	100
70	7950	10	10
60	4350		

Figure 2 is an attempt to make the equal loudness contours more meaningful by translating phons into L.U.'s (×10) using the table published by the American Standards Association (60) and then plotting the contours so that each successive contour going up the scale represents a doubling of the loudness.

Two basic methods have been employed in measuring the loudness of sounds. The *first method* or older of the two is the binaural matching method with or without an off-set receiver. In this procedure a normal observer is provided with a receiver being driven through a calibrated attenuator by some kind of oscillating circuit. While listening to the sound to be measured with one ear he matches the loudness in the other ear by adjusting the attenuator. The attenuator has usually been calibrated in terms of db. above the threshold for a 1000 cycle tone. A variation of this method employed in some of the earlier work was the use of a phonograph record of a warbling tone. The gain control was calibrated in decibels above the hearing threshold.

Some question has been raised regarding the advisability of using a receiver, offset from the ear a centimeter or more, for the purpose of mixing the known and unknown sounds. Those favoring the use of an offset argue that comparisons are made more accurate because both the known and unknown sounds are matched in a single ear. On the other hand others (5) insist the offset has a baffle effect on the incoming waves from the unknown source distorting their loudness. Practiced observers following a rigid variant of the method of limits familiar to psychophysicists have in this manner measured complex sounds with a fair degree of reliability. "The spread of the mean final settings normally does not exceed ±2 db. at 100 db. level, ±4 db. at 80 db., and ±6 db. at 50 db." (5). Some experience using a receiver without an offset while

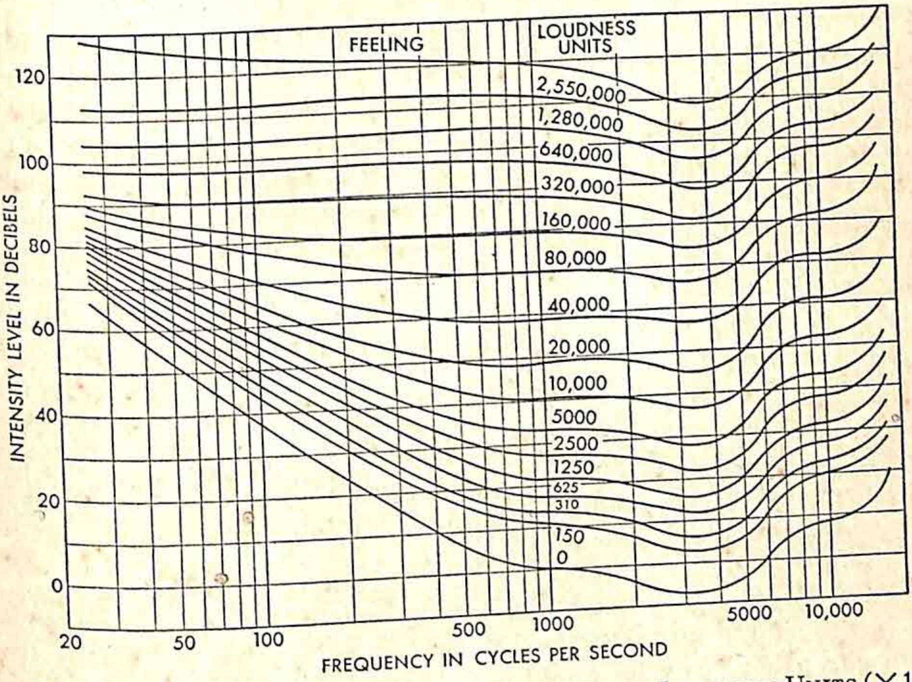


FIGURE 2*. EQUAL LOUDNESS CONTOURS EXPRESSED IN LOUDNESS UNITS (×10)

matching the loudness of an unknown noise indicates that the position of the receiver is highly critical. Slight pressure on it or a small shift with respect to the meatus produces large changes in the apparent loudness of the calibrated sound.

---

* In preparing Figure 2, loudness units were multiplied by 10. To reduce, neglect the final zero, i.e. 2,550,000=255,000 L.U.'s and so on.

Using a different arrangement Geiger and Abbott (20) presented twenty observers with sounds from a variety of sources, (buzzer, rattler, vacuum cleaner, bell, etc.) by means of a microphone, amplifier, and loud speaker which were to be matched in loudness by a 1000 cycle tone introduced also through a loud speaker. The mean difference for all sounds and all observers between the sound level as measured by a meter and the matched tone was 2.4 db., the meter reading being between 76 and 40 db. Generally the observers matched the unknown sound level with a 1000 cycle tone at a higher energy value than the sound level meter indicated. It is significant that Geiger and Abbott concluded, "This was a very severe test indeed, and the meter measurements were completely confirmed."

The *second method* employed in measuring sound levels is the objective method in which a microphone is placed in the sound field and the electrical output of the microphone is amplified. An adjustable weighting network may be interposed between this amplified current and the indicating meter usually calibrated in decibels. The meter actually indicates A.C. power which is proportional to the rate of reception of sound energy at the microphone. The weighting network is adjustable so that the response of the instrument will approximate the response of the ear to the sound energies falling on it.

Referring to Figure 1, it can be seen that between 30 and 1500 cycles at 100 db. energy intensity, loudness remains practically constant. However, if the energy intensity level is reduced to 40 db. the frequencies below 90 cycles would not be heard and those above 90 would vary from just perceptible to 40 phons in loudness. Consequently a noise meter must be adjustable so that at high intensities it gives a reasonably constant output for all frequencies and with low intensities it should give little or no response at low frequencies, but should increase its output as the frequency is raised to approximately 2000 cycles and then decrease as the frequency exceeds this level. Since it is not at present practically feasible to build a portable meter with networks permitting response curves similiar to each of the loudness contours most commercial sound-level meters are provided with three such networks adaptable to levels approximating 40, 70, and 100 db. in over all intensity.

Under special conditions where the noise to be measured is made up of a number of harmonic frequencies as is the case with the noise from large transformers the objective method gives readings that may be as much as 30 db. *lower* than the aural comparison method (5). The reason for this lies in two related facts. First, frequencies close to, but not exactly equal to the harmonics of a given fundamental exert a masking effect on that fundamental thereby reducing the loudness which would otherwise be present. Second, when a number of harmonic components are present in the noise being measured their energies add to, rather than subtract from the loudness in a manner not provided for in the sound-level meters now available (56). The meters commercially available compensate for the masking effect of the near-harmonics but do not sum the sound energies of the harmonics as the ear does. Geiger and Abbott noted in the report referred to above (20) that in cases where the sound was composed of a large number of "component musical notes, well separated in frequency and of nearly the same loudness" the individual ratings of loudness were considerably higher than the meter readings.

Another limitation of sound level meters traceable to the microphone design is the wide tolerance permitted in their response characteristics by the American Standards Association. These are such that a 90 db. reading may have an error of plus or minus 3 db. On the surface such an error may appear small yet if one should take a low reading instrument to measure the sound level in a given area of approximately 90 db., then return later after acoustical treatment with a high reading instrument, no change would be registered. Subjectively, however, unpracticed observers might judge the noise after acoustical treatment to be half as loud as before. Indeed, a change of 6 db. at that level is almost equal to doubling or halving the loudness as determined by a number of independent investigators using trained observers under careful laboratory conditions (19, 24, 26). In many field measurements of noise these errors are small relative to the fluctuation in sound energy from moment to moment and place to place.

It becomes clear that the aural comparison method is superior to the objective method is some instances provided that competent observers are employed, and the test tone is properly calibrated. The known variation in individual observers are of the same order as the allowable tolerances in microphone response characteristics. In addition, however, the sound-level meters even if accurate in their response to the sound energy of the field do not register the non-masking effects of harmonic components in a sound spectrum. Undoubtedly one of the chief reasons for the wide use of sound level meters rather than the aural comparison method lies in the fact that the meter readings are repeatable with a high degree of reliability. Furthermore, it is difficult if not impossible for an observer to match sounds varying greatly in intensity from moment to moment. The aural comparison method does show differences in loudness when the same observer makes repeated observations of the same sound. This is especially true of inexperienced observers and leads investigators after a brief trial of the method to discard it. On the other hand the work of Geiger and Firestone (21), Ham and Parkinson (23), and Fletcher and Munson (17) showed that practiced observers following carefully a standard procedure are remarkably reliable in their estimations of loudness of relatively steady tones.

It is possible to compute the loudness of a complex noise by the use of a wave analyzer which records the magnitude of sound energy in each of several wave-frequency bands. The noise spectrum thus determined provides the basic data for computing the loudness using formulae developed by Fletcher and Munson (17). The method is cumbersome and yields results only slightly different than the mean value of several observers using a properly calibrated test sound.

It becomes clear that both methods of measuring noise levels requires more than an ability to set up a microphone, flick switches and read dials. The meter readings must be interpreted in the light of the nature of the sound spectrum and the known tolerances of the meter itself. On the other hand the aural comparisons must likewise be conducted and interpreted by one familiar with the standard manner of listening and the known sources and magnitude of errors. In the hands of competent investigators either method will not give a measure of high precision but will provide a basis for an informed estimate of loudness.

Effects on Production

One may find in textbooks references to a number of reports indicating that production is markedly increased with a reduction in noise. Several of these emanate from an anonymous article (62) in which references to the original reports are lacking. It is not possible, therefore, to evaluate the claims that "lowering the noise level in a telephone-exchange room from 50 db. to 35 db. resulted in a forty-two per cent reduction in errors and a three per cent reduction in cost per message" (4). Or that "moving the assembly department of a temperature-regulator company from next to the boiler room to a quieter location reduced faulty assemblies from seventy-five per cent to seven per cent of those turned in for inspection" (4).

Only one field study in this country has been completed with sufficient control and covering a sufficient time to place confidence in the results. This study has not been formally published in detail although it has been referred to frequently (**32**, **59**). The Aetna Life Insurance Company in December 1928 installed sound absorbing materials in an office occupied by typists, clerical checkers, punchcard and comptometer operators. Bonus figures were available reflecting the efficiency of employees for a year prior to installation. This "efficiency" refers to the allowed time for a given amount of work as determined by time studies, divided by the actual time taken. The average difference in the bonus based on efficiency of fifteen clerks taken at semi-monthly intervals for the year prior to, and a year after acoustical treatment was 9.2% in favor of the quieter condition. At no time during the "quiet" year did the bonus go below the level of the first year. No other changes were introduced into the working conditions during these two years. The sound level as measured by a Western Electric 3-A audiometer using the loudness matching method was 35 "sensation units"* after acoustical treatment.

A control check on these findings was made after one year by covering the sound absorbing materials with gypsum board. The sound level was then 41 "sensation units." The bonus efficiency immediately dropped to an intermediate value between the two previous years, but within two months was approximately equal to the level of the "quiet" year. At the time the bonus reached this high level in the presence of the greater noise there was a reduction in work, only the more efficient employees being retained from then on, which probably accounted for the high average efficiency. Comparable, but less clear-cut results were noted among the punchcard, and comptometer operators. In these latter groups machine breakdowns and other uncontrolled factors complicated an interpretation of the data.

The available data provides no basis for the calculation of statistical tests of reliability of the differences noted. In spite of this fact it is likely that the advantage noted in favor of the quieter condition is not a matter of chance.

The influence of noise in school children and grade school work was investigated by Obata (**39**). The noises were a variety of phonograph records supplemented by a mechanical rattler varying between 60 and 85 db. With each type of sound working speed was more affected than accuracy in arithmetic and symbol cancellation. The adverse effects duced a chance superiority for the experimental groups. It is regrettable that the investigators did not interchange the experimental and control groups.

---

* This is a unit roughly equal to a decibel and used prior to 1936.

The influence of suggestion was not controlled. How important a factor suggestion may be can only be surmised. Baker (3) demonstrated with college students working for 5 minute periods under noise and quiet conditions that output can be lowered or raised depending upon subtle suggestions incorporated in the pre-experimental instructions. We have no way of knowing what suggestions were given workers in the Weston and Adams studies. However, since neither the work, nor the length of the work spells in Baker's study are comparable to the industrial situation there is no reason to expect suggestion would be a major factor influencing personal efficiency.

Finally, the variability of percentage improvement in personal efficiency from month to month ranging from 31.42% to 3.42% suggests powerful uncontrolled factors influencing the differences between the groups.

There remains, none the less, the fact of a consistent superiority of output under sound reduction. The authors' conclusions that "excessive noise is a factor having an important effect upon industrial efficiency" and that "tolerance is not acquired with lapse of time" are not irrefutably established, but the weight of evidence is in their favor.

A study by Kornhauser (28) attempted to relate the output of typists to noise in a regular office situation. The results were equivocal and complicated by the existence of other distractors in the quieter location.

Using only two subjects, presumably the authors, Vernon and Warner (54), reported that performance on an arithmetic test, a purely routine assembly job, and reading difficult material were not significantly affected by noise after a short adjustment period during which there were adverse effects. On the other hand judgments of the magnitude of subjective disturbance indicated less rapid adjustment to the noise. These conclusions were based not only upon observations under laboratory conditions but included tests made in the presence of factory and office noises.

In reviewing these reports dealing with production and accuracy of work one is impressed with the methodological difficulties of conducting meaningful investigations "in the field." Other conditions beyond experimental control, such as temperature, machine breakdowns, or changes in quality of raw materials may obscure whatever benefits noise reduction may contribute. It seems clear that a thorough study of industrial production as contrasted with office work has yet to be done under positive control of all factors, covering work periods of sufficient length with noise typical of industrial situations. Furthermore, the factor of suggestion in the studies so far published has received little attention possibly because it is difficult to control.

INFLUENCES ON VITAL PROCESSES

The effect of noise on blood pressure, respiration rate, metabolism, muscular tension, electrical skin resistance, digestive processes and similar functions has attracted some experimental investigation. Among the most widely quoted of these is the report by Laird (29) who measured the metabolic rate of four typists during a half hour of resting and while typing a standard letter over and over again for two hours per day throughout a four-week period. During the first and last of these weeks the walls of the testroom were bare. During the second and third weeks sound absorbing material was applied to walls and ceiling reducing the noise produced mechanically from 50 "sensation units" to 40 "sensation units."

On the average the metabolic rate was fifty-one per cent higher while working compared with the resting rate during the quieter weeks, and was seventy-one per cent higher during the noisy weeks. A suggestion that the noisy phase was more fatiguing is found in the data showing the average time for the last five letters of the two-hour spell was seven seconds less than for the first five letters in the quiet phase, while the comparable time was five seconds more in the noisier condition. However, the fast typists improved in speed when the noise was reduced, while the slow typists showed little or no change in overall speed.

No day-to-day data were published by Laird, a fact that has been criticized by Harmon (24), who found a very rapid adjustment to noise, using metabolic rate as the principal indicating function and mental arithmetic as the work. Such a task has a much lower muscular component then typing and could not be expected to affect the metabolic rate to a marked degree. It is important to note that on the initial days of noise Harmon found a thirty-seven per cent increase in metabolic rate. During the same days the average number of errors per day was 12 for quiet and 24.3 when working under noise.

Prior to Laird's study Morgan (38) found evidence of similar extra effort in the presence of noise. His subjects engaged in a form of translating letters into numbers by pressing appropriate keys. Respiration was recorded as well as the pressure exerted upon the keys. The noise came from a variety of buzzers, bells and phonograph records. Noise and quiet periods alternated during the experimental sessions; the ten minute noise period following thirty minutes of quiet. Noise was responsible for greater key pressure being exerted and greater articulation on the part of the subjects—a device spontaneously adopted in an apparent attempt to overcome the distracting effects of the noise. The time per response increased significantly at the beginning of each noise period but subsequently decreased to a point where output was more rapid than during quiet.

Vernon and Warner (**54**), Freeman (**19**), and Ford (**18**) generally confirm the findings of Morgan and Laird so far as an increase in energy expenditure in the presence of noise is concerned.

Davis (**9**) measured muscle tension in the presence of noise by means of action potentials. His eighteen subjects were required to do nothing except to sit and listen to the noise presented in a series of two-minute intervals separated by two minutes of quiet. Marked increases in tension appeared on the first experimental day at the onset of each noise period. On subsequent days the change became less and less marked so that by the fifth day tension was virtually unchanged when noise was presented.

The effect of noise on digestive processes was examined by Smith and Laird (**49**) whose four subjects swallowed a rubber balloon which was later inflated in the stomach. Peristaltic contractions were thus recorded by an appropriate pneumatic system operating a writing point on a kymograph. Noise of 80 db. for ten minutes following a twenty minute quiet period produced a thirty-seven per cent decrease in the number of contractions. The recovery period following the noise showed no consistent pattern or change in stomach action. The effects using a 60 db. noise were less marked but of the same nature. The rate at which saliva and gastric juices flowed was also decreased in the presence of noise. These changes would presumably have an adverse effect upon the total digestive process.

Although details of the experiment have not been published in full, Luckiesh (**33**) reported that a "demonstration visual test" was performed six per cent more quickly in a quiet room than in the presence of the hum from a motor-generator set.

Other observations in the presence of noise of high intensities and short duration have demonstrated increased blood pressure (**35**) and increased cranial pressure (**26**). In the later case, however, one wonders to what degree the sound pressure itself may have affected the recording apparatus.

Davis (**10**) reported a marked decrease in skin-resistance at the onset of a five-minute noise period. However, the function gradually returned to normal and reached a value before the end of the noise period approximating the value obtained with a control group sitting quietly for ten minutes. The noise level is not reported.

Even though no experimental studies have been reported dealing with effect of noise on emotional control or emotional stability, the reports above indicate physiological changes of the kind associated with emotional disturbances. Many of the changes noted in response to noise are similar to, if not identical with those produced by stimuli exciting fear, anger, and general emotional tension. Throughout much of the published literature (**15**) are testimonials that cannot be dismissed lightly indicating that high noise levels are the cause of "nervous exhaustion" emotional instability and related conditions.

ADAPTATION TO NOISE

In some of the studies previously cited, notably those by Harmon (**24**), Morgan (**38**), Ford (**18**), Davis (**10**), Pollock and Bartlett (**44**), emphasis has been given to the process of adaptation or adjustment to noise. Morgan found an initial decrement in output followed in some instances by greater-than-normal production. Harmon reported an initial increase in metabolic rate followed by smaller and smaller increments on succeeding days at the onset of noise. Pollock and Bartlett's subjects reported they could hardly bear the noise in the early experimental sessions yet at the end they did not notice the noise. These reports do not fit with the long term studies by Laird, Aetna Life Insurance Company, Weston and Adams, which show a net deleterious effect of noise. It is possible that the differences in results can be attributed, in part, to the factor of suggestion as already mentioned in the previous section. On the other hand it is not reasonable to expect that a higher-than-average efficiency index could be maintained without a break over a full year of operation in a group of fifteen clerks by suggestion alone, as the report of the Aetna study shows. It is also true that the studies showing an adjustment to noise such that the detrimental effects are reduced or eliminated cover, for the most part, short noise periods. We may hazard the guess that if the noise were continued over longer periods comparable to those experienced in industry the adjustment mechanism compensating quickly for the initial ill effects may break down and the net effect of noise even in later studies might have been adverse. In line with this hypothesis is Freeman's (**19**) idea that the added effort noted in the presence of the initial noise is not just a rise in total energy expenditure but comes from a shift in the pattern of supporting processes, not directly concerned with the principal functions (typing movement, etc.) under observation. With continued exposure to the noise this irradiation of excess tension becomes less prominent and a more economical concentration of supporting processes reasserts itself.

FACTORS AFFECTING THE ANNOYANCE OF NOISE

Throughout the studies mentioned above are frequent references to the erratic manner in which noise acts as a distractor. Only a few studies have been reported which were aimed at discovering the features of the noise which contributed to annoyance. Using a paired-comparison method with pure tones at 50 db., Laird and Coye (31) found the annoyance of tones to be ranked in the following descending order: 8192, 4096, 64, 2048; with 128, 256, and 1024 of minimum and equal annoyance. Thus, other things being equal, high tones and extremely low tones are judged more annoying then those in the middle range. The effect of intensity was examined by matching in annoyance a 256 cycle tone at various intensities with other pure tones adjustable in loudness. The equal annoyance contours thus established below 500 cycles followed roughly the equal loudness contours shown in Figure 1. Above that frequency the equal annoyance contours tended to drop away from the equal loudness contours, indicating that low tones of high intensity are as annoying as high tones of relatively low intensity.

The observations of acoustical engineers (40) testify to the importance of high frequencies as they affect the reported comfort of treated areas. In one case two adjacent offices were acoustically treated, but with different materials both having the same noise level and similar sound sources. The absorption coefficient at 4096 cycles for Office A was 0.45 and for Office B was 0.82. B was rated by its occupants as comfortable while those working in A complained that the typewriter noise was not properly quieted. Replacing the lower coefficient material by the other type corrected the complaint, in spite of the fact that the overall Noise Reduction Coefficients for the two materials were approximately the same.

Interrupted noise or discontinuous tones have been generally found to be more annoying than steady noises. This was true in the reports by Pollack and Bartlett (44), and Laird (24).

Furthermore Finch and Culler (13) found greater auditory destruction in dogs subjected to high intensities when the tones were interrupted. Even the reports of Morgan (38), Davis (10), Freeman (19), Harmon (24), and others who have used relatively short noise exposure periods would lead one to expect that discontinuous noise would be the more detrimental. In all of these reports the initial onset of noise produced the greatest adverse effect. Hence the discontinuous noise would be somewhat comparable to a rapid series of noise periods each of which demands a new adjustment with initial phases of maximum cost to the individual.

A number of writers have referred to a miscellaneous list of factors that influence the annoyance of noise, none of which have been subjected to systematic investigation. Among these factors are the unexpectedness of noise (42), the spreading effect and reverberation (47) the degree to which the noise is unnecessary or indicates malfunctioning of equipment (2).

Of primary interest to psychologists are the individual differences in noise tolerance. Numerous independent observers in the field (47) as well as nearly all laboratory investigations agree in reporting wide differences in the degree to which persons are affected by the same noises. Related to this factor is the wide variation in response by a given person on different occasions. Shifts of attitude, motivation, and attention result in widely different types of performance under the same external noise conditions. No systematic attack has been made on this problem to discover what personality feature or other psychological factors are correlated with noise tolerance.

STIMULATION DEAFNESS

Kemp (25) has recently reviewed the studies dealing with stimulation deafness. It is, therefore, unnecessary to review that area except to reemphasize the necessity of interpreting the reported incidence of alleged occupational deafness in the light of the normal decrement in acuity with age. The most complete information dealing with incidence of hearing losses in the general population is that gathered by the Bell Telephone Laboratories at the New York and San Francisco Fairs in 1939 (50). In spite of this precaution which prevents the acceptance of many reports at their face value, there is little doubt that extended exposure to intense noise will produce both a temporary and permanent loss in auditory acuity.

Much of the experimental work in this area has been done on animals in an effort to throw light on the mechanism of hearing. The results have shown in general that the frequencies suffering the greatest impairment are not correlated with the predominant exposure frequencies, although the degree of impairment is related to the duration and intensity of the exposure noises (8, 43, 46, 55).

Although numerous reports are available purporting to show that workers in high noise levels have a greater-than-average incidence of hearing loss (22, 34, 35, 36, 37, 40, 45, 52, 53), the devices for measuring hearing have varied widely making any comparison between occupations hazardous. One survey (61) has been published of nine plants in New York State conducted in a standard manner using a Western Electric 3-A audiometer. The report concludes that the greatest percentage of deaf people occurs in plants of greatest noise. However, it must be pointed out that not all employees in the plants of greatest noise were tested. It is possible that a selective factor operated to exaggerate the proportion of deaf people tested in those plants.

There appears to be no reasonable doubt that high noise levels can and do produce hearing defects especially after long exposure. Still unanswered are the questions, what levels can be tolerated, and for how long without incurring a defect?

CONCLUSION

In spite of a wide spread interest in noise abatement relatively few facts have been well established. Popular literature not covered in this review abounds in emotional outbursts against the baneful effects of noise. Public support has been enlisted for noise abatement campaigns on the uncritical acceptance of the assumption that noise because it is annoying must be harmful. The available scientifically controlled studies are not in complete agreement but tend to show ill effects in output, speed of work or vital processes. Although a considerable degree of adaptation takes place, the evidence suggests that it is seldom complete. Marked individual differences in susceptibility to the ill effects of noise have been noted but no reported attempts have been made to correlate these differences with other facets of personality. The factors determining annoyance have not been subjected to thorough analysis. Stimulation deafness is an unquestioned result of exposure to loud noises for long periods but its extent and the critical levels of noise necessary to produce it in humans have not been clearly established. In summary, it is clear that there are many circumstances under which noise detracts from efficiency and well-being. Under what circumstances noise is deleterious and for what kinds of people are questions for further fruitful research.

BIBLIOGRAPHY

1. Abbott, E. J. Quieting substation equipment. *Elec. Eng.*, 1935, **54**, 20–26.
2. Abbott, E. J. Sound measurements in industry. *J. Ap. Phys.*, 1939, **10**, 526–531.
3. Baker, K. H. Pre-experimental set in distraction experiments. *J. appl. Psychol.*, 1937, **16**, 471–488.
4. Berrien, F. K. *Practical psychology.* New York: Macmillan Co., 1944.
5. Churcher, B. G. Noise measurements for engineering purposes. *Elec. Eng.* 1935, **54**, 55–65.
6. Churcher, B. G. & King, A. J. Analysis and measurement of the noise emitted by machinery. *J. Inst. Elec. Eng.*, 1930, **68**, No. 397.
7. Dashiell, J. F. *General psychology.* Boston: Houghton Mifflin Co., 1937.
8. Davis, H., Derbyshire, A. J., et al. Experimental stimulation deafness. *Science*, 1935, **81**, 101.
9. Davis, R. C. The muscular tension reflex and two of its modifying conditions. *Indiana Univ. Pub.* Science Series, 1935. No. 3.
10. Davis, R. C. Electrical skin resistance before, during, and after a period of noise stimulation. *J. exp. Psychol.*, 1932, **15**, 108–117.
11. Dockeray, F. C. *Psychology.* New York: Prentice Hall, 1942.
12. Dorman, P. Annoying noises of our offices. *Northwest Med.*, 1932, **31**, 317–321.
13. Finch, F. & Culler, E. Effects of protracted exposure to a loud tone. *Science*, 1934, **80**, 41–42.
14. Firestone, F. A. The technique of sound measurements. *J. Soc. Auto. Eng.*, 1926, **19**, 461.
15. Flaiz, J. W. & Wolf, E. G. The effect of noise on aviation personnel. *Contact*, Pensacola, 1943, **3**, 160–167.
16. Fletcher, H. Loudness, masking and their relations to hearing and noise measurement. *J. acoust. Soc. Amer.*, 1938, 9, 275–293.
17. Fletcher, H. & Munson, W. A. Loudness, its definition, measurement and calculation. *J. acoust. Soc. Amer.*, 1933, 5, 82–108.
18. Ford, A. Attention-automatization: an investigation of the transitional nature of mind. *Amer. J. Psychol.*, 1929, **41**, 1–32.
19. Freeman, G. L. Changes in tension pattern and total energy expenditure during adaptation to distracting stimuli. *Amer. J. Psychol.*, 1939, **52**, 354–360.
20. Geiger, P. H. & Abbott, E. J. Sound measurements versus observers' judgment of loudness. *Elec. Eng.*, 1933, **52**, 809–813.
21. Geiger, P. H. & Firestone, F. A. Estimation of fractional loudness. *J. acoust. Soc. Amer.*, 1933, **5**, 25–30.
22. Gilbert, D. J. Influence of industrial noises. *J. industr. Hyg.*, 1921, **3**, 264–275.
23. Ham, L. B. & Parkinson, J. S. Loudness and intensity relations. *J. acoust. Soc. Amer.*, 1932, **3**, 511–534.
24. Harmon, F. L. The effects of noise upon certain psychological and physiological processes. *Arch. Psychol.*, 1933, No. 147.
25. Kemp, E. H. A critical review of experiments on the problem of stimulation deafness. *Psychol. Bull.*, 1935, **32**, 325–342.
26. Kennedy, F. Fatigue and noise in industry. *N. Y. St. J. Med.*, 1936, **36**, 1927.
27. Kingsbury, A. Direct comparison of the loudness of pure tones. *Phys. Rev.*, 1927, **29**, 588–600.
28. Kornhauser, A. W. The effects of noise on office output. *Industr. Psychol.*, 1927, **2**, 621–622.

29. Laird, D. A. Influence of noise on production and fatigue, as related to pitch, sensation level and steadiness of the noise. *J. appl. Psychol.*, 1933, **17**, 320–330.
30. Laird, D. A. Experiments in the physiological cost of noise. *J. Nat. Inst. Ind. Psychol.*, 1929, **4**, 251–258.
31. Laird, D. A. & Coye K. Psychological measurements of annoyance as related to pitch and loudness. *J. acoust. Soc. Amer.*, 1929, **1**, 158–
32. Lindahl, R. Noise in industry. *Ind. Med.*, 1938, **7**, 664–669.
33. Luckiesh, M. Visual efficiency in quiet and noisy work places. *Elec. W.*, 1931, **98**, 472.
34. McCord, C. P. The heart problem of workers. *Assn. of life insurance med. dir. of Am.*, 1931, **17**, 52.
35. McCord, C. P., Teal, E. E., & Witheridge, W. N. Noise and its effect on human beings. *J. Amer. med. Ass.*, 1938, **110**, 1553–1560.
36. McKelvie, W. B. Weaver's deafness. *J. Laryng, Rhin. and Otol.*, 1933, **48**, 607.
37. Mollison, W. M. Ear and occupation. *Guys Hospital Reports*, 1936, **86**, Series 4, **16**, 185–190.
38. Morgan, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35.
39. Obata, Juichi, Morita, S., Hirose K., & Motsumoto, H. Effects of noise on human efficiency, *J. acoust. Soc. Amer.*, 1934, **5**, 255–261.
40. Parkinson, J. S. & Jack, W. A. Reexamination of the noise reduction coefficient. *J. acoust. Soc. Amer.*, 1941, **13**, 163–169.
41. Pedley, F. G. Incidence of occupational deafness and methods for its prevention. *Safety Eng.*, 1930, **60**, 237.
42. Poffenberger, A. T. & Rounds, G. H. Unpublished Manuscript.
43. Pofroff, N. F. On the changes in the auditory apparatus of white mice caused by noises in the metal trades: preliminary report from the all-Russian congress of otolaryngologists held at Moscow in 1927. *J. industr. Hyg.*, and cited by McCord, C. P., Teel, E. E. and Witheridge, W. N. Noise and its effect on human beings. *J. Amer. med. Ass.*, 1938, **110**, 1555.
44. Pollock, K. G. & Bartlett, F. C. Psychological experiments on the effect of noise. *Indus. Health Res. Bd.*, 1932, Rep. 65.
45. Roger, L. R. Noise-deafness: a review of recent experimental work and a clinical investigation into the effect of loud noises upon the labyrinth in boiler-makers. *J. Laryng., Rhin., and Otol.*, 1915, **30**, 91–105.
46. Rosenblith, W. A. Industrial noises and industrial deafness. *J. acoust. Soc. Amer.*, 1942, **13**, 220–225.
47. Sabine, H. J. & Wilson, R. A. The application of sound absorption to factory noise problems. *J. acoust. Soc. Amer.*, 1943, **15**, 27–31.
48. Scheidt, V. P. The effects of various modifications on the worker. *Kalenda*, 1937, **10**, 3–4.
49. Smith, E. L. & Laird, D. A. The loudness of auditory stimuli which affect stomach contraction in healthy human beings. *J. acoust. Soc. Amer.*, 1930, **2**, 94–98.
50. Steinberg, J. C., Montgomery, H. C. & Gardner, M. B. Results of the World's Fair hearing tests. *J. acoust. Soc. Amer.*, 1940, **12**, 291–301.
51. Stevens, S. S. & Davis, H. *Hearing*. New York: Wiley & Sons, 1938.
52. Swann, C. C. The effect of noise on hearing. *Internat. J. Med. & Surg.*, 1933, **46**, 314–315.
53. Vadala, A. J. Effects of gun explosion on the ear and hearing mechanisms. *Mil. Surgeon*, 1930, **66**, 810–822.
54. Vernon, H. M. & Warner, G. G. Ob-

jective and subjective tests for noise. *Personnel J.*, 1932–33, **11**, 141–149.

55. Weaver, E. G. & Smith, K. R. The problem of stimulation deafness. I. Cochlear impairment as a function of tonal frequency. *J. exp. Psychol.*, 1944, **34**, 239–245.
56. Wegel, R. L. & Lane, C. E. Auditory masking of one pure tone by another. *Phys. Rev.*, 1924, **23**, 266–285.
57. Weston, H. C. & Adams, S. The effects of noise on the performance of weavers. *Indust. Health Research Bd.*, 1932, Rep. 65.
58. Weston, H. C. & Adams, S. The performance of weavers under varying conditions of noise. *Indust. Health Research Bd.*, 1935, Rep. 70.
59. Wilson, A. Better concentration reduces employee turnover 47 percent, adjustment of noise factor increased general efficiency 8.8 percent. *Bankers Mo.*, 1942, **59**, 254–255.
60. *Acoustical Terminology. Booklet 224.1*—1942. *Noise measurement Booklet 224.2*—1942. American Standards Association.
61. *Effect of noise on hearing of industrial workers.* [Anon.] N. Y. State Dept. Labor, Spec. Bull. 166, 1930.
62. *Effect of noise upon efficiency.* [Anon.] *Monthly Labor Review,* 1930, **30**. 1199–1208.

STUDIES IN TIME PERCEPTION

A. R. GILLILAND, JERRY HOFELD AND GORDON ECKSTRAND
Northwestern University

The study of time takes various forms and presents problems to investigators in many fields. For example, the nature of time is of special interest to the philosopher. The accurate measurement of time is important for the physicist and the astronomer. In a somewhat related sense, time is a very practical problem for the daily worker who punches a time clock, the athlete who competes in games and the aviator in his maneuvers. The perception of time is a subject of special interest to the psychologist.

In this paper we shall be interested in time perception and time estimation and shall be concerned with such problems as the nature of time only in so far as it is necessary for a definition of our problem. In this quest we will not be greatly aided by such definitions of time as that proposed by Kant who said that time is "a universal form of intellect" or by Leibnitz who described it as "the obscure and confused picture of the grounds which determine the order of succession."

The physicist's definition of time, as the measurement of the movement of the earth through space, is more helpful but it is not entirely satisfactory for our purposes. Neither is the somewhat naive operational definition of the psychologist who says that time is what the clock measures, much more helpful. This is true for the double reason that time existed and no doubt was perceived before the advent of clocks and secondly, because when we ask what it is that the clock measures we are back to our original question.

Such considerations in the absence of any more constructive material leads us to the tentative description of time as our perception of one of the two essential characteristics of movement. One of the essential characteristics is space and the other is time. This brings out the close relationship between time and space. Such an explanation smacks strongly of the theory of relativity and fits nicely with Gestalt theory in psychology. However, this need not necessarily constitute a criticism of the definition. Neither does it inevitably commit us to a purely Gestalt explanation of time since such time-space relations may be assumed to be the result of our experiences rather than nativistically given.

If we wish a more specific definition of time it may be described as the perception or probably better stated as the judgment of the amount of separation of two units or limits as greater or smaller, in terms of some unit such as seconds. This judgment is reached from an estimation of how long it would take to accomplish so much or so many things such as to walk a block, listen to so much or to feel rested or fatigued by so much work.

While many theories have been proposed and many experiments conducted on the subject of time perception or time estimation* as yet there is no generally accepted view as to how we perceive or estimate time. Studies in time estimation might well be classified under several related but somewhat different problems such as,—

1. The estimation of relatively short time intervals. Most of the experimental work in time has dealt with the estimation of time between two signals up to a few minutes in length. In most cases the estimator has been forewarned that an estimation was to be made.

2. The estimation of time when the subject is not expecting to make such estimation. For example, if the reader was suddenly asked to estimate how long had been spent in reading this paper up to this point, we do not know whether the same methods of estimation would be used as in case he had expected to make this estimation.

3. It is also not certain whether the same methods are used in estimating and in reproducing time intervals. In estimating time a unit of estimate such as a second is necessary. In reproducing time, a unit is at least not so necessary.

4. The perception of longer intervals involving hours or days may be similar or may be different from estimating shorter periods of time. Certainly some of the same factors are involved but some new ones may also be involved.

5. There also is the question of how lower animals perceive time. Do they use the same methods as humans? At least there certainly is little "mental content" used in their estimations.

6. Then there is the problem of estimating the time of day or night. This at least in modern society involves direct reference to some instrument of time measurement such as the clock. In primitive society the sun or moon was, no doubt, generally used as the reference point. That our modern methods are only a refinement of this method is indicated by our usual methods of indicating time by so many hours, minutes and seconds since noon or midnight.

In the following sections of this paper we shall review the principal recent studies that have dealt with the problems listed above. Since, research has not proceeded as logically as the outline, this review will follow the various lines of research.

Since there have been excellent reviews of the literature on time estimation in a series of articles by Dunlap (13) from 1911 to 1916 and a later review by Weber (56) in 1933, this article will in general be limited to the literature since the later date. There will, however, be occasional violations of this principle (1) in order to emphasize some point of view developed during the earlier period and (2) in order to summarize studies in time perception in animals. This later topic was not included in the earlier reviews and it may well have an important bearing on research in time perception.

---

* The phrases perception of time and estimation of time are here used somewhat interchangeably. In a sense they are not strictly synonymous. Time estimation always implies a kind of quantification not necessarily included in the time perception.

CUES FOR TIME PERCEPTION

One of the important problems in time perception is that of the cues used. In fact, most of the experimental studies in time perception can be related more or less directly to this problem. In the closely allied field of space perception the principal cues as well as their relative importance, as Carr (8) has so well pointed out, have been very carefully investigated. Although many studies have been conducted, no such list of cues has been discovered for time perception. In fact, almost the opposite is true. That is, we are hardly sure of any of the cues that are used in time perception.

Several of the earlier studies reviewed by Weber (56) investigated the relationship between certain physiological processes such as breathing, fatigue or digestive processes and time estimation. Some of these studies found evidence for such relationships but the evidence was never conclusive.

In some recent studies in this field Rosenzweig and Koht (45) compared time estimation and what they called "need tension." They defined need tension as a state of strain. They found that during periods of greater need tension there was more underestimation that when the tension was not so great.

A somewhat similar point of view was presented earlier by Francois (17, 18) and later by Hoaglund (30) who believed that time is mediated by internal temperature fluctuations. For example, Francois found that when the internal temperature was increased by high frequency currents, the tapping rate of the subject was increased. From such kinds of studies these two writers believe we have some kind of internal time clock for the perception of time.

Kawasima (37) had subjects estimate time while moving their arms in ten degree circles. Overestimation occurred when the action was easy and underestimation when it was difficult.

Shaefer and Gilliland (48) systematically varied (1) pulse rate (2) heart work (3) blood pressure (4) breathing rate and (5) lung work, in order to determine the effects of each on the estimation of time. No relationships were found between any of these variables and time estimation. Even when all these changes occurred together producing a state of great physiological activity they seemed to have no influence on the direction of error or accuracy of time estimation.

Berman (**5**) investigated the relation of time estimation to what he called satiation. A stylus maze was used to produce satiation. When the subjects were satiated, 87 per cent of his subjects underestimated the time required for them to become satiated while 52 per cent of the non-satiated group overestimated the time required to reach their criterion of learning.

Jasper and Shogass (**34**) found that the alpha rhythm of the brain was in no way related to conscious estimates of time.

Certain drugs and disease affect the estimation of time. It is common knowledge that the drunk man is generally disoriented as to time as well as space.

Favilli (**15**) states that under mescal intoxication time estimation is very poor. Forty-five minutes was estimated as 15. But sometimes the estimates are abnormally short and sometimes abnormally long. Favilli explains this by suggesting that in the intoxicated state the shortening of the "field of consciousness" results in a sort of contemplation of the present instant which lacks a frame of reference. This seems to be a way of saying that the subject is always living for the present and lacks the usual perspectives for estimating time. If this be true, mescal intoxication may not have a greatly different effect upon time estimation than alcoholic intoxication. Marijuana also seems to produce a somewhat similar effect. It is sometimes smoked in cigarettes by orchestra leaders and drummers with the idea that it assists them in rapid timing. Whether it does or not seems not to have been investigated experimentally.

Sterzinger (**49, 50**) gave subjects several different drugs for the purpose of determining their effects on time estimation. After taking quinine subjects habitually underestimated 5 minute intervals. Alcohol produced too low an estimate for periods less than 25 minutes and too high for longer periods. Caffeine and thyroxin gave inconsistent results. Sterzinger came to the conclusion that time estimation depended upon metabolic rate.

FILLED AND UNFILLED TIME

One of the early topics to receive experimental study in the field of time estimation was the effect of filled and unfilled time. That is, what effect does the type or amount of mental content or physical activity have upon its apparent length? Although the literature of this topic is extensive the results, as will be seen from this survey, are by no means consistent. Some results seem to indicate as William James (35) suggested that filled time while passing is perceived as longer than it really is. On the other hand many studies have obtained the opposite results. One explanation for such discrepancies is the fact that it is difficult to determine whether a time interval is filled or empty. The interval may be filled with a large amount of visual, auditory, or other type of external change. However, these may have little or no attention value for the subject who is estimating the time. On the other hand, the time interval may be relatively quiet so far as external stimuli are concerned and still be richly filled with mental content for the subject. For example, he may be thinking, "now one second, now two seconds, three seconds, etc." Many of the seeming inconsistencies in the results for different investigators are due to factors such as those just enumerated.

Much of the early experimental work on filled versus unfilled time was summarized by Triplett in 1931 (52).

Later studies by Helm (29) considered subjective factors such as attention as important in time estimation. Harton (26, 27, 28) found that activities of varying difficulty influenced time estimation. Time spent in making difficult discriminations was estimated as less than equal periods spent in making easy discriminations. In working at tasks in which the subjects were encouraged and felt they were being successful, the time was estimated as shorter than when they were discouraged and felt they were being unsuccessful.

There could hardly be more inconsistency in results than those found in the various studies in estimating filled and unfilled time. Does filled time appear longer or shorter than unfilled time? Does William James' dictum that filled time seems shorter in passing and longer in retrospect hold true? As already suggested, no single answer can be given to this problem except that time estimation depends upon other factors than whether the time is filled or unfilled. In so far as filled time may more often appear longer than unfilled time, it is largely a matter of the subjective filling of the time.

Recent studies by Abbe (1, 2) investigated the effect of increasing the amount of physical space upon time perception. Two blinking lights were used to delimit the time interval. The distance between these lights were varied but the time between their flashes was left constant. Abbe reports that time intervals were reported as longer when the space between the lights was greater. In a related experiment he also found the converse to be true, that is, if space intervals were given in a longer interval of time the space between the lights was judged as longer than when given in a shorter time. Tashiro (51) in a similar experiment in which he could arbitrarily vary both the rate of tapping and the space between which the taps were made, concludes that the rate of tapping built up a gestalt with the result that the meaning of each tap was lost irrespective of the tapping speed. However, the tempo of the tapping which was determined solely by the spatial distance between the points struck in the taps was more powerful in producing the effect of gestalt movement than the rate of tapping. Rudert-Kotte (46) holds that binocular parallax, so important in space perception, may give clues to time perception through the space concept as well as through the slight time variants due to the fact that we see through two eyes instead of one.

While such studies as those just enumerated do not prove conclusively any theory of time perception, they give us material which must be taken into consideration in the formulation of any theory and they strongly indicate that there is a close relationship between the amount of mental content and the estimation of time. Space is directly related to time in that when space is extended this gives the impression of greater mental content. Yet this no doubt is entirely too simple an explanation of time and will need further elaboration as we proceed.

INDIVIDUAL DIFFERENCES IN TIME PERCEPTION

Three problems relating to time perception will be discussed in this section. One of these is the determination of the indifference point in time estimation. Another deals with the Weber-Fechner law and the other with sex differences in time estimation.

Most of the work on the indifference point in time estimation occurred before 1933 and was reviewed in Weber's (56) article. The results may be summarized by saying that different investigators have obtained very different results.

Similar inconsistencies have been obtained in determining whether Weber's law holds in time estimation. Some evidence indicated that errors are proportional to the length of the interval and time studies show no such relationship. In more recent studies Gilliland (20) using over 300 subjects with a total of about 5000 estimations ranging in length from 4 to 27 seconds found evidence favoring the Weber-Fechner law. But in another study by Gilliland and Humphreys (21) the per cent of errors was found to be significantly greater for the shorter than for the longer intervals. In this study the intervals ranged from 9 to 180 seconds. It seems likely that the difference in results in these later studies was due to the wider range of time intervals in the later study. It seems probable that since time is not mediated by any one sense but through several different senses no single constant could be established in the Weber-Fechner formula. This may be especially true when the intervals vary greatly in length.

In the earlier studies of sex differences women were generally found to be poorer than men in time estimation. But in later studies such as those by Harton (**26, 27, 28**) and Gilliland and Humphreys (**21**) no such differences were found. It seems likely that the modern woman is called upon to make time estimates as often as men. Since time perceptions are primarily dependent upon learning the sexes probably do not differ in ability to estimate time.

TIME PERCEPTION IN LOWER ANIMALS

Some knowledge of how time is perceived may be gained from a study of time perception in lower animals. Because we do not believe that the lower animals, at least, use mental imagery to anything like the extent that humans do, any theory that is based upon mental content alone tends to be discredited.

The principal experimental studies of time perception in lower animals may be classified under the general headings of (1) delayed choice response, (2) delayed conditioned responses, and (3) time perception.

There is considerable doubt whether delayed choice responses should be considered under the heading of time perception. The delay of the choice response may not be due to an appreciation of time interval by the animal making the delayed response. Nevertheless, a brief account of such experiments will be included here since the length of the delay is one of the important factors in such reactions.

Hunter (**33**) using a multiple choice apparatus, measured the time that could elapse between the presentation of a light before the correct choice box and the release of the animal which would result in the animal going to the correct box. In a three choice box, Hunter found that for different rats the time could vary from one to ten seconds but the average time was not much more than one second. For dogs the time averaged about ten seconds and for raccoons it averaged about 20 seconds.

Walton (**54**) using a different kind of choice box obtained delays up to one minute in dogs. Warden and Warner (**55**) made tests on the noted dog Fellow and obtained delays of about 20 seconds between spoken commands and their execution. Yarbrough (**59**) obtained from two to four seconds delay in cats on a three compartmental and 16 to 18 seconds on a two compartment maze. But Corvan (**9**) obtained delays of 30 seconds with a Persian cat. In contrast with the previous studies Honzik (**31**) secured delays up to 45 seconds with rats.

The great variance of these results no doubt partly illustrates individual differences in animals but it also illustrates differences in methods used and apparatus. Because of these differences in results no very definite generalizations can be made other than to point out that the higher the animal in the scale, the more likely that the delay can be longer. It is also true that in the lower species the delay is maintained in terms of bodily posture whereas higher animals can change posture and still react correctly after a delay.

Several authors have studied time responses in invertebrates. Verlaine (53) taught wasps to discriminate between alleys of different lengths. He thought this showed a "concept of time." Braunschmid (7) made studies on time responses of fishes while Grabensberger (23) worked with ants and Sterzinger (49, 50) with ants, bees, and wasps. Richter (44) and Johnson (36) have studied the diurnal rhythms of rodents. Johnson found that this rhythm developed in wild mice born and reared in utter darkness. Beling (4) working with bees found that he could condition them to return for feeding any time of day or night.

Most of this work shows that some kind of relationship exists between diurnal rhythm, metabolic rate, and some sort of a time sense. But this sense of time is no doubt very different from time perceptions in humans.

Kuroda (40) by means of delayed conditioning found that rats were quite accurate up to about 20 minutes and that cats could be given delayed conditioning up to about 30 minutes. Anderson (3) experimented on time discrimination in the white rat. He used a multiple choice maze and confined the rat for different periods of time. Rats were able to discriminate between periods of 10 and 20 seconds. Kuo (39) constructed a maze with four pathways to the food box. Some were longer than others and in one the rat was confined for 20 seconds. The rats learned to go by the shortest route and avoided the one in which they were confined first of all. These results were further substantiated by an experiment by Sams and Tolman (47). Of course, it must be realized that other factors than time perception affect the order of elimination of pathways in such experiments.

Woodrow (58) taught two monkeys to differentiate between temporal intervals of 1.5 and 4.5 seconds. They were taught to reach for food upon the raising of a screen after the longer but not after the shorter interval. The monkeys became 90 per cent correct for these time intervals and 75 per cent correct in distinguishing between 1.5 and 2.18 seconds. Metfessel and Bobbitt (41) found that canaries could be taught to discriminate and produce rhythms according to a prearranged pattern. They acquired an accuracy with an average deviation of only .01 seconds.

LEARNING IN TIME PERCEPTION

While some writers claim that the perception of time like other kinds of perception is largely dependent upon certain innate factors, probably no one would deny that learning plays a part in the process. The important questions are how does learning take place and how much of a part does it play. As indicated in the early part of this paper there are probably several different problems relating to time estimation. If so, then there are several problems related to learning in time perception.

The effect of different cultures on the development of time concepts is stressed in two papers. Kloos (38) states that the Persians stress the future while the Chinese stress duration, and the Indians stress the stagnation of time. Rodulescu-Motru (43) believes that the cultural effect amounts to a process of evolution whereby time is evolved through periodic changes in nature, in the body, and in culture. Thus he claims that part of the process of time perception is genetically inherited.

In the field of the experimental study of the time perception in children Elkine (14) reports that in children from ten to sixteen years of age there were large errors in time estimation. He found average errors of 82 per cent for short intervals up to 4.1 seconds and 57 per cent for longer intervals for the younger children. In sixteen year old children the error was 32 per cent for short intervals and 30 per cent for longer intervals. Gilliland and Humphreys (21) found average errors ranging from 32 to 47.5 per cent in fifth grade children as contrasted with errors of 15 to 31 per cent for adults. No doubt the greater accuracy in children in the latter study as contrasted with Elkine's study is due to the fact that Elkine studied Russian children from the home of workers whereas Gilliland and Humphreys studied American children from better homes. Such children have more occasion to be influenced by time and thereby learn the meaning of time intervals at an earlier age than the Russian children.

Several studies have investigated what sense departments are best suited for perceiving time. Goodfellow (22) found that, considering audition as 100 per cent, vision gave 65 per cent and the tactile sense 34. Other sense departments were even lower than any of these. Gault and Goodfellow (19) made two studies; one concerned with discrimination efficiency and the other with reproductive efficiency. In both the auditory sense was most efficient, vision second and touch was third. Forbenius (16) had subjects wake from sleep at previously arranged times. He reports that they became able to do so within an accuracy of about five minutes. Wirth (57) taught subjects to reproduce given time intervals. Dudycha and Dudycha (11) asked 185 subjects to estimate the amount of time it would take them to perform six different tasks.

Gilliland (20) taught a group of 13 subjects to estimate time by counting. After each estimate they were told the length of the interval. In 25 to 50 practices these subjects were able to reduce their average error from 20 to 25 per cent to 5 to 10 per cent. A similar group with an equal amount of practice but without counting did not improve in their estimating.

THEORIES OF TIME PERCEPTION

Despite the large amount of disagreement in results in studies on time, many theories have been proposed to explain the perception of time. Most of the older theories postulated some kind of a time sense. The theory presented by Francois (17, 18) and Hoaglund (30) assuming the existence of some kind of an internal chemical time clock comes under this general heading. But later studies have not confirmed this point of view.

Psychoanalysts generally attribute time perception to some phase of the self. Hugenholz (32) explains time as follows: "The covert human time-form can exist only when there is a consciousness of self." Bonaparte (6) tells us that the child in whom the unconscious dominates has only a vague concept of time as in day dreaming and other forms of imagining. Dooley (10) explains time as an "Ego-defense properly placed in that class of defenses comprised of attachments to real objects in counterdistinction to narcissistic attachments." Just what is meant by such explanations and how convincing they are, is left to the reader.

Obendorf (42) explains time as "primarily dependent upon the reality with which recurrent physiological reflexes are registered. It becomes distorted in the absence of the sense of reality and in cases of profound interference with the perception of reality time sense may cease to exist. Another factor which is strongly influential upon time perception is the presence or absence of purpose, for without purpose the value of time greatly diminishes." Sterzinger (49, 50) as has already been stated believes time perceptions are based upon body metabolism, when metabolism is high, time passes quickly, when it is low, time drags. This bears a close relationship to the theory of Francois and Hoaglund and is essentially the theory proposed by du Nouy (12). Dunlap (13) stresses the relationship between time perception and bodily rhythm.

A number of writers emphasize the Gestalt point of view in time perception. Weber (56) stresses the relationship between time and space. In a series of experiments by a number of *Gestalt* writers the close relationship between time and space is emphasized. This relationship is expressed by the formula "phenomenal velocity equals phenomenal space divided by phenomenal time or $v = s/t$." This is further emphasized in the statement "these variations in the flow of phenomenal time are not occasional examples which may be explained as illusions, but that they are continuous and are conditioned by almost any change in the structure of the entire field of movement."

While the evidence is either too conflicting or inconclusive to prove or disprove any theory of time perception, an analysis of the data lends support to certain general principles about how time is perceived. May we here sketch briefly some of these.

As has already been stated the evidence does not support a theory of any direct sensory basis for time perception. Unlike space perception and the estimation of distance time is less directly dependent on sense experiences as such. In the case of lower animals a discrimination of longer or shorter intervals and even some vague notion of time are, no doubt, mediated by physiological processes and body rhythm.

In children this vague awareness of time is further amplified by learning units of time such as a second or hour. This knowledge is gained either directly or indirectly from timepieces. The significance of these intervals are interpreted in terms of body tensions and rhythm. As the child gains further experience it learns more about evaluating how much activity can occur in any given time. The sensory cues can come from any sense modality and the resulting activity may be physiological or mental.

In the case of the rat or dog in the delayed conditioned response it is certainly mostly physical. In the case of the reader in estimating how long it has taken him to read this paper his estimate is probably based upon his estimate or knowledge of how many pages he has read and at what rate he reads a page. Likewise if a person knows that he ordinarily walks about five miles per hour when he has walked what he thinks is about three miles he will estimate that he has been walking for 40 minutes. If a subject is not paying attention to what is going on during the interval or if he has no good idea of the amount that has taken place or of units for measuring the interval, his estimates will be poor. Training as shown in learning the normal rate of one's counting during an interval will soon make a person able to estimate time very accurately. Without some kind of definable content during the interval, estimation remains poor.

In retrospect estimation depends upon the memory of events occurring within any interval. If the interval is filled with many events it will seem long in recall. If it is uneventful it will be remembered as short. Similarly in immediate estimation the estimation will depend upon the amount that has occurred or seemed to have occurred during the interval. It is for this reason that when we are doing interesting things time is generally underestimated. In waiting or boredom time passes slowly. However, there is always a correction that the estimator may make. If a person knows that time is underestimated during pleasant happenings and overestimated in boredom he may correct or often overcorrect for these influences. It is, no doubt, in large measure because of these facts that the literature on filled and unfilled time is so hard to interpret.

According to this theory time estimation is not nearly so directly given from sense data as in space perception and estimation. Time estimation partakes more of the nature of a judgment rather than a perceptual process. The fact that in space perception the cues remain for further examination whereas in time perception after the event it is always in retrospect, helps to make time estimation more of a process of judgment than of perception.

What has thus far been presented as a brief theory of time perception applies particularly to short intervals. Possibly longer intervals are estimated in somewhat the same way but more often with greater reference to external events such as the position of the sun and familiar events such as the amount of traffic on the street as in estimating the time of day. Of course, when a timepiece is available it will be used.

In an earlier part of this paper six general fields of research in time perception were indicated. These are areas in which further research could add much to our knowledge about time. In addition to these general fields of research much insight might be gained about the nature of time from a further study of time estimation in animals. The evolution and improvement of time perception both in children and adults would also, no doubt, throw further light on this subject. Such studies would have both practical value in showing how to train persons in time estimation and would extend our knowledge about the psychological nature of time.

BIBLIOGRAPHY

1. Abbe, M. The spatial effect upon the perception of time. *Jap. J. exp. Psychol.*, 1936, **3**, 1–52.
2. Abbe, M. The temporal effect upon the perception of time. *Jap. J. exp. Psychol.*, 1937, **4**, 83–93.
3. Anderson, A. C. Time discrimination in the white rat. *J. comp. Psychol.*, 1932, **13**, 27–55.
4. Beling, I. Über das Zeitgedachtnis bei Bienen. *Z. vergl. Physiol.*, 1929, **9**, 259–338.
5. Berman, A. The relation of time estimation to satiation. *J. exp. Psychol.*, 1939, **25**, 281–293.
6. Bonaparte, M. L'inconscient et le temps. *Rev. franc. Psychanal.*, 1939, **11**, 61–103.
7. Braunschmid, M. Zeitsinn bei Fischen. *Blat. Aquarien., u. Terrarienk*, 1930, **41**, 232–233.
8. Carr, H. A. *An introduction to space perception.* New York: Longmans, 1935.
9. Cowan, E. A. An experimental testing of the ability of the cat to make delayed response and to maintain a given response toward a varying stimulus. *J. comp. Psychol.*, 1923, **3**, 1–9.
10. Dooley, L. The concept of time in defense of ego integrity. *Psychiatry*, 1941, **4**, 13–23.
11. Dudycha, G. J., & Dudycha, M. M. The estimation of performance time in simple tasks. *J. appl. Psychol.*, 1938, **22**, 79–86.
12. du Nouy, P. L. *Biological time.* New York: Macmillan, 1937.
13. Dunlap, K. Rhythm and time. *Psychol. Bull.*, 1911, **8**, 230–242; 1912, **9**, 197–199; 1914, **11**, 169–171; 1916 **13**, 206–208.
14. Elkine, D. De l'orientation de l'enfant d'age Schdaire dans les relations temporelles. *J. de. Psychol.*, 1928, **25**, 425–429.
15. Favilli, M. La percezione del tempo nell'-ebbrezza mescalinica. *Rass. Studi psichiat.*, 1937, **26**, 455–462.
16. Forbenius, K. Über die Zietliche Orientierung im Schlaf und einige Aufwachphanomene. *Z. Psychol.*, **103**, 100–110.
17. Francois, M. Contribution a l'etude du sens du temps. La temperature interne, comme facteur de varition de l'appreciation subjective des dureés. *L'anée Psychol.*, 1927, **28**, 186–204.
18. Francois, M. Influence de la temperature interne sur notre appreciation du temps. *C. R. Soc. Biol., Paris*, 1928, **108**, 201–203.
19. Gault, R. H. & Goodfellow, L. D. An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. *J. gen. Psychol.*, 1938, **18**, 41–47.
20. Gilliland, A. R. Some factors in estimating short time intervals. *J. exp. Psychol.*, 1940, **27**, 243–255.
21. Gilliland, A. R. & Humphreys, D. W. Age, sex, method, and interval as variable in time estimation. *J. gen. Psychol.*, 1943, **6**, 123–130.
22. Goodfellow, L. D. An empirical comparison on audition, vision, and touch in the discrimination of short intervals of time. *Amer. J. Psychol.*, 1934, **46**, 243–258.
23. Grabensberger, W. Untersuchungen uber das Zeitgedachtnis der ameisen und termiten. *Z. vergl. Physiol.*, 1933, **20**, 1–55.
24. Groethuysen, B. De quelques aspects du temps. Notes pour une phenomendogie du recit. *Rech. Phil.*, 1935–36, **5**, 139–195.
25. Harrison, M. L. Development of time in children. *Elem. Sch. J.*, 1934, **34**, 507–514.
26. Harton, J. J. The influence of the dif-

ficulty of activity on the estimation of time. *J. exp. Psychol.*, 1938, **23**, 270–287.

27. Harton, J. J. The influence of the difficulty of activity on the estimation of time. *J. exp. Psychol.*, 1938, **23**, 428–433.
28. Harton, J. J. The relation of time estimates to the actual time. *J. gen. Psychol.*, 1939, **21**, 219–224.
29. Helm, W. Die Beeinflussung des Vergleichswertes einer zweiten teilweise gleichzeitigen Zeitstrecke. *Arch. ges. Psychol.*, 1937, **98**, 490–556.
30. Hoagland, H. The physiological control of judgments of duration: evidence for a chemical clock. *J. gen. Psychol.*, 1933, **9**, 267–287.
31. Honzik, C. H. Delayed reactions in rats. *J. comp. Psychol.*, 1931, **15**, 307–318.
32. Hugenholtz, P. T. Over tijd en tijdsvormen. *Ned. Tijdschr. Psychol.*, 1938, **6**, 143–161.
33. Hunter, W. S. The delayed reaction in animals and children. *Behav. Monogr.*, 1913, **2**, No. 6.
34. Jasper, H. & Shagass, C. Conscious time judgments related to conditioned time intervals and voluntary control of the alpha rhythm. *J. exp. Psychol.*, 1941, **28**, 503–508.
35. James, Wm. *Principles of psychology*, New York: Henry Holt, 1890, Vol. 1, Ch. 15.
36. Johnson, M. S. Activity and distribution of certain wild mice in relation to biotic communities. *J. Mammal.*, 1926, **7**, 245–277.
37. Kawasima, S. The influence of time intervals upon the perception of arm motion. *Jap. J. Psychol.*, 1937, **12**, 270–289.
38. Kloos, G. Storungen des Zeiterlebens in der endogenen Depression. *Nervenarzt*, 1938, **11**, 225–244.
39. Kuo, Z. Y. Genesis of the cat's responses toward the rat. *J. comp. Psychol.*, 1930, **11**, 1–36.
40. Kurdoda, R. Time estimation of longer intervals in the white rat. *Acta Psychol. Keijo.*, 1936, **2**, 155–159.
41. Metfessel, M. & Bobbitt, J. M. Pitch and time accuracy of birds. *Psychol. Bull.*, 1934, **31**, 602–603.
42. Oberndorf, C. P. Time—its relation to reality and purpose. *Psychoanal. Rev.*, 1941, **28**, 139–155.
43. Radulescu-Notru, C. Noi cercetari asupra timpului. *An. Inst. Psicol.*, 1937, **4**, 1–34.
44. Richter, C. P. A behavioristic study of the activity of the rat. *Comp. Psychol., Monogr.* 1922, **1**, No. 2, 1–55.
45. Rosenzweig, S. & Koht, A. C. The experience of duration affected by need-tension. *J. exp. Psychol.*, 1933, **16**, 745–775.
46. Rudert-Kotte, R. Die Binokulare zeitparallaxe unter typologischem gesichtspunkte. *Z. angew Psychol.*, 1937, **53**, 251–260.
47. Sams, C. F. & Tolman, E. C. Time discrimination in white rats. *J. comp. Psychol.*, 1925, **5**, 255–263
48. Schaefer, V. G. & Gilliland, A. R. The relation of time estimation to certain physiological changes. *J. exp. Psychol.*, 1938, **23**, 545–552.
49. Sterzinger, O. Chemopsychologische untersuchungen über den Zeitsinn. *Z. Psychol.*, 1935, **134**, 100–131.
50. Sterzinger, O. Neue chemopsychologische untersuchungen über den menschlichen Zeitsinn. *Z. Psychol.*, 1938, **143**, 391–406.
51. Tasiro, T. On the temporal factors in the Gestalt of movement. *Jap. J. Psychol.*, 1937, **12**, 290–308.
52. Triplett, Dorothy. The relation between physical pattern and reproduction of short time intervals. *Psychometr. Monogr.*, 1931, **41**, No. 87, 202–265.
53. Verlaine, L. L'instinct et l'intelligence chez les hymenopteres. IX.

La notion du temps. *An. et Bull. Soc. Entom.*, de Belg. 1929, **69**, 115–125.

54. Walton, A. C. The influence of diverting stimuli during delayed reaction in dogs. *J. ani. Behav.*, 1915, **5**, 259–291.

55. Warden, C. J. & Warner, L. H. The sensory capacities and intelligence of dogs with a report on the ability of the noted dog "Fellow" to respond to verbal stimuli. *Quart. Rev. Biol.*, 1928, **3**, 1–28.

56. Weber, A. O. Estimation of time. *Psychol. Bull.*, 1933, **30**, 233–252.

57. Wirth, W. Die unmittelbare Teilung einer gegebenen Zeitstrecke. *Amer. J. Psychol.*, 1937, **50**, 79–96.

58. Woodrow, H. Temporal discriminations in the monkey. *J. comp. Psychol.*, 1928, **8**, 395, 427.

59. Yarbrough, J. W. The delayed reaction with sound and light in cats. *J. Anim. Behav.*, 1917, **7**, 87–110.

BOOK REVIEWS

STEINER, LEE R. *Where do people take their troubles?* Boston: Houghton Mifflin, 1945. Pp. xiii+265.

What do we psychologists do when our personal problems become so severe we are ready to jump off a tall building or take other drastic action? More adequate adjustive behavior involves consulting with a respected colleague or a psychiatrist. Mrs. Steiner is concerned with what the great gullible public does with its problems. Her book indicates that a surprisingly large number of otherwise fairly intelligent people consult such purveyors of psychotherapy as newspaper columnists, radio "consultants," marriage or "get acquainted" bureaus, religious faddists, spiritualists, palmists, and numerologists. These "therapists" compensate for their lack of formal training by assuming an air of great self-confidence and something approaching omniscience.

The book begins with a useful discussion of "What is a psychologist?" It is pointed out that there are recognized courses of training preparing people to be psychiatrists, psychiatric social workers, psychoanalysts, and clinical psychologists. Examination of telephone directories and contacts with license bureaus and health departments in big cities showed an almost complete lack of regulation of people dispensing psychotherapy as long as they avoided hypnosis and the treatment of serious mental disorders.

What sorts of therapy and therapist appear in this wide open field? It is easy to discount "graduates" from the "College of Divine Metaphysics." This institution confers a "Doctor Degree in Psychology" on the basis of two courses, costing fifty dollars each. A surprising thing is the large number of similar institutions giving "degrees." A common characteristic of most of the "therapists" discussed by Mrs. Steiner is their lack even of one of these dubious degrees—one simply adds to one's name a set of letters like Ms.D. or Ps.D. After all, why bother to go to school when your practice is based on common sense, of which you clearly have more than your share?

As a circulation-builder the newspaper columnist-consultant is probably useful to the publisher. However, the advice given may often be harmful and is rarely helpful. Mrs. Steiner sent a copy of the same letter to a number of leading columnists, briefly outlining a mother's problem with a disobedient child. One advocated sending the boy to a special school, and several others proposed stern discipline of one sort or another. Similarly, advice based on inadequate evidence is freely given on a number of radio programs listened to by millions. Both types of "advisors" have in common a certain incompetence and an engrossing interest in the commercial aspects of their ventures.

The alleged therapist who deals with an office clientele is most obviously the profiteer on others' emotional problems. Mrs. Steiner devotes the greater part of her book to a carefully documented report of her experiences with these people and their "patients." We can visit with her the "Human Audit Bureau" where we are given vocational guidance on the basis of our brain development as measured with special calipers. If we need to find companionship of a sort, we can always contact a marriage broker or some introduction service. After all, "Why let lonesomeness rob you of romance, happiness, and comfort?" What shall we do if we are really interested in religion but are not satisfied with the regular recognized churches because they don't seem to solve our personal difficulties? This problem is "easy." There are many local "churches" to take care of us—would we prefer spiritualism mixed with a little hypnosis, or stick with the "I Am" system? Perhaps we are prejudiced in favor of something like palmistry, graphology, astrology, numerology, or tarot; many practitioners of these occult sciences parade through the chapters self-confidently if somewhat denuded of respectability.

Mrs. Steiner is not one of the smoothest of writers. This reader's sensation was often one of bumping along from one section to another with little help in transition. The overwhelming mass of apparently scrupulously recorded and reported experiences making up the book's documentation probably contributes to this feeling of disjointedness. However, the subject material of the book will be of great intrinsic interest to any professionally-minded psychologist, and a careful reading will repay anyone with even the glimmerings of a "social conscience."

A review would be incomplete without at least a mention of some of the book's obvious possibilities for professional use. One could:

1. Recommend it to his more intelligent students in elementary or clinical psychology courses as supplementary reading material.

2. Pass it along as encouragement or reminder to committees in professional organizations engaged in formulation and acceptance of standards to be met by practicing applied psychologists.

3. Use it to strike lightly those psychologists who have obstructed the setting up of such professional standards.

4. Bring it to the attention of legislators and point out similar conditions existing in their own constituencies, hoping that legislative support may thus be obtained for regulation of people unwarrantedly calling themselves psychologists.

5. Write letters quoting its conclusions to newspapers and periodicals accepting advertisements from individuals and organizations similar to those discussed.

6. Pass along the information from it in any ethical, practical way to the public which has so often been victimized.

7. Use of an occasional rereading of sections of the book to reawaken one's

conscience and remind oneself that such abuses do exist, and that it is the professional psychologist's duty to eliminate them.

Robert Ammons.

State University of Iowa.

Kaplan, Oscar J. (Ed.), *Mental disorders in later life.* Stanford, Calif.: Stanford Univ. Press, 1945. Pp. vii+436.

The disparity between the available research on the early segment of development (childhood), and the late (senescence), has long been obvious. The paucity of factual material available on mental decline has no doubt been disquieting to the teacher of developmental psychology who strove to present consistently the entire sequence. Particularly is this void conspicuous when aging is viewed as the ultimate consequence of maturation and integration. Moreover, this period of the individual's growth assumes increased importance since more persons now live to be old, thus steadily increasing the ratio of old people to those in other age brackets.

The editor modestly presents the seventeen chapters as a collection of essays and not "a full summary of the vast literature already available in this field." Although the title and general orientation of the volume emphasizes mental disorders, several of the chapters are good, up-to-date reviews of typical changes with aging.

In these days of specialism it is heartening to find men of diverse scientific backgrounds collaborating on the treatment of phenomena of a limited field, even though this cooperative effort is additive rather than integrative. A brief, general introduction of the volume by a psychiatrist (Karl M. Bowman) is followed by a review of later life disorders by a statistician (Horatio M. Pollock). Well documented systematic articles on the physiological (Nathan W. Shock) and psychological (Harold E. Jones and Oscar J. Kaplan) aspects of normal and abnormal age changes follow. To complete the survey of older age processes the newer and highly significant sociological (H. Warren Dunham) and nutritional (L. Erwin Wexberg) approaches are added.

Chapters on the *Neuroses* (Norman Cameron), the *Involutional Psychoses* (Eugene Davidoff), the *Presenile Dementias* (George A. Jervis), the *Senile Psychoses and Psychoses with Cerebral Arteriosclerosis* (David Rothschild), the *Aged Subnormal* (Oscar J. Kaplan), and the *Toxic Delirious Reactions* (G. Wilse Robinson, Jr.), outline the clinical disorders of this age period with an emphasis on the recent literature. *Psychotherapy* (Fred V. Rockwell) and *Mental Hygiene* (Nolan D. C. Lewis) each merit a chapter in addition to sections on treatment in the clinical chapters. A systematic chapter ends the collection. It is written by a gerontologist (Edward J. Stieglitz) who has a broad perspective on the phenomena of aging and its unsolved problems. The chapter does not, however, specifically articulate with those preceding it. Sandwiched between the clinical types is a novel chapter describing a study in one institution of *Older mental patients after long hospitalization* (Eugenia Hanfmann).

Like most volumes of this kind, the style, general outlines, extent of organization of facts and completeness of coverage of the materials vary greatly. The editor does not state the degree to which each writer was cognizant of the other contributors' outlines or manuscripts. However, the overlapping and repetitions that occur do not detract from the book.

The contributors differ in the extent to which they deal in mere description and in hypothetical explanation. Some approach their phenomena from an atomic and some from a multiphasic standpoint. Wexberg, in treating nutrition, finds it necessary to discuss employability of the aged. In a discussion of neuroses, Cameron generalizes more as to the basic nature of psychological changes than do the authors of the chapter devoted to that subject. Shock integrates his detailed, scholarly chapter, based upon 281 titles, and attempts to explain impairment of mental functions by the hypothesis that aging is attributed to progressive loss in the homeostatic capacities of the organism. Dunham, with less data, presents a framework to interpret the interplay of culture upon the biological factors.

This factual volume is a distinct contribution to abnormal and developmental psychology. Dr. Kaplan is to be commended for his broad perspective of the problem, his appeal to specialists for data, and his excellent choice of contributors. However, with full cognizance of the enormity of the task and the potential vulnerability of the author, one does regret that a chapter has not been added to attempt to integrate all of the contributions.

University of Missouri. FRED MCKINNEY.

SHERMAN, M. *Intelligence and its deviations.* New York: Ronald Press, 1945. Pp. x+286.

The first part of this brief text describes the nature of intelligence, mental growth, and the relation of mental growth to physical and environmental factors. There follow brief discussions of intelligence in delinquency and psychosis, but most of the discussion of deviations concerns mental deficiency in its association with neurological disorders. There are brief sketches of intelligence tests, of genetic problems in intelligence, and of intellectual superiority.

The treatment of the topics is very uneven. The author seems at home in the medical discussion of such conditions as cretinism, microcephaly, birth injury, epilepsy and hydrocephalus, and the information, presented rather technically, will be of interest to the psychologist. There is a chapter on the adjustment of the mental defective which, while short, surveys a number of studies. As a general text, however, this book is not satisfactory.

In a general discussion of intelligence and its deviations, which it is the author's stated intention to provide, one would expect a well-rounded development of the many facets that the subject possesses. Yet, in this very brief book, there is a very meagre treatment of most of the topics, and some important subjects are not mentioned at all. One feels that Professor Sherman wished to present the medical and neurological aspects of certain conditions which occur in some defectives, and then added material deemed necessary to prevent the book's having a rather limited and specialized appeal.

There is a very brief and inadequate discussion of the rationale of intelligence testing, a knowledge of which is indispensable to the systematic understanding of the field. There is very little mention of qualitative features of test performance, test patterns, problems of rapport, differences between diagnostic groups (some of which are never mentioned), or of many other significant and important topics. From the book, one would conclude that the only really significant deviation from normal intelligence is in the direction of deficiency. Seven pages are devoted to the intellectually gifted! In the chapter on intelligence and delinquency, the major problem raised is whether the average intelligence of delinquent groups is below that of other groups. Only eight pages are devoted to intellectual changes in psychoses other than the psychoses of mental defectives, and little of the extensive literature in this area is cited. Few of the basic problems studied are made known. It is a sterile and rather futile picture of the study of intellectual deviation that is revealed in these pages. None of the studies of intellectual function of such investigators as Goldstein, Hanfmann and Kasanin, N. Cameron, J. Hunt (to mention but a few) is mentioned, and the whole problem of test performance in relation to lobectomy is ignored.

The emphasis in presentation is well illustrated in the chapter on epilepsy, to which 19 pages are devoted. Three pages are given to intellectual changes (deterioration). The nature of epilepsy, its age of onset, personality changes, and its pathology are described, and one derives a reasonably adequate picture of the disease. But this is a book about *intelligence* and *its* deviations.

The discussions of general topics, such as theories, growth, and the like, are poorly organized and will hardly leave the student with a clear idea as to what the problems are or what has been done in their solution. Professor Sherman feels, for example, that IQ differences found in identical twins reared separately are an indication of "achievement levels" dependent upon different environments. This, of course, is the crux of the matter. But he then argues that such results cannot evidence the influence of environment on intelligence, because tests are not culture free (253). Is the concept of intelligence itself culture free?

There are a glossary of terms, mostly medical, and a bibliography of over 300 titles. It is perhaps significant that but seven titles bear publication dates later than 1939.

CHARLES N. COFER.

George Washington University.

BRENNAN, ROBERT E. *History of psychology, from the standpoint of a Thomist.* New York: Macmillan, 1945. Pp. xvi+277.

The author presents and evaluates the history of psychology from the point of view of Thomas Acquinas, whose views in turn were based upon the writings of Aristotle. Nearly half of this condensed history is devoted to the antecedents, philosophic and otherwise, of psychology as a separate field. The remainder of the book is devoted to an accurate but very brief presentation of the important events, research, and personalities in the development of modern psychology. The following is evidence of the brevity of the presentation: Wundt, and his achievements, are presented in 4 pages, although there are references to his name or work on 18 other pages; William James is given 3 pages and 4 other references; while Freud is given 13 pages and 9 other references.

Historical sequence is carefully followed *within* the development of a topic, but the omission of dates limits the possibility of temporal cross reference from one topic to another. This is compensated for in part by the inclusion of a bibliographical index which gives the years of birth and death as well as listing the chief publication of the individual by title.

The author's viewpoint is epitomized in the following paragraph,

> What we need today, as Acquinas would contend, is really less of psychology and more of anthropology, using the term "anthropology" in its traditional meaning to signify the study of man, as man, not as a concatenation of reflexes, or a sum of perceptual configurations, or a series of imaginal processes or a complexus of instinctive responses. Such things are simply isolated events in the history of human nature, and they have no meaning except in relation to the whole nature. Further, the study of man, as man, means the study of man as a besouled organism, or as a creature composed of matter and spirit, whose operations fall within the dimensions of scientific analysis, but whose fundamental nature is the proper study of philosophy. (257)

The research man will probably find the material of this book too condensed and the references too limited to serve his interests. Many an academic man will object to the philosophical and religious emphases as being too extensive for his effective use in the classroom. However, considering the religious and philosophic background of the author, most psychological readers will be gratified to observe the emphasis placed on and the respect shown toward the *established scientific observations* in all fields of psychological investigation.

WILSON MCTEER.

Wayne University.

TRAXLER, A. E. *Techniques of guidance: tests, records, and counseling in a guidance program.* New York: Harper, 1945. Pp. xiv+394.

This comprehensive survey of methods and procedures for use in guiding and counseling high school students is written for the benefit of high school teachers and administrators. The book includes (xiii) "much detailed explanation concerning tests and other instruments of evaluation and a large number of illustrative record forms," written to advise school personnel who may have had little psychological training. Detailed and pertinent suggestions on a variety of associated administrative problems are also provided. Such a book is to be evaluated not as a scientific treatise but as a guide to the proper application of psychological techniques and procedures.

Apart from a brief discussion of occupational opportunities early in the book and an equally brief treatment of "adjustment" problems at the end, the primary concern is with problems of distribution or placement of students within the educational system. The book is organized around the individual cumulative record, which is to include information obtained periodically for each student in the following ten areas: home background, health, out-of-school experiences, plans for the future, school history and record of class work, mental ability or academic aptitude, special aptitudes, achievement and growth in different fields of study, educational and vocational interests, and personality. Information in the first four areas can be obtained from each individual by interview or questionnaire, and school records accumulate more or less automatically.

Information in the last five areas involves the use of tests and observational procedures, and it is here that a detailed explanation of methods of obtaining and applying such information is needed. A minimal explanation of the applicability of a measure in a guidance situation should include a statement of the validity of the measure in predicting scholastic, vocational, or other performance, so presented that a statistically unsophisticated person could grasp the actuarial nature of the problem. This we are nowhere given.

The chapter on the appraisal of aptitudes describes or mentions a number of standard tests of intelligence or general academic aptitude suitable for use with high school students. A carefully prepared annotated list of thirteen such tests, which is presented at the end of the chapter, gives for each test the author(s), publisher, cost per copy, mention of the component parts, and reliability coefficients. For only one of the thirteen tests is validity in relation to scholastic performance mentioned at all. Since tests of this type do have substantial validity for many purposes, and considerably less validity for many others, it would seem that a more intensive consideration and explanation of this subject would lead to a more discriminating use of such tests.

Under "special aptitudes," similar information is given for tests of art, music, science, mechanical, clerical, mathematics, and foreign language aptitudes. Counselors are warned that the label of a specific aptitude test does not always indicate accurately what the test actually measures, but methods of choosing the most effective of several similarly named tests are not provided.

A chapter on the evaluation of achievement presents an exhaustive annotated list of achievement tests in all scholastic areas.

In presenting a catalogue of personality tests, questionnaires and interest inventories, the author is carefully critical of some of the procedures reviewed; others are simply mentioned or described briefly. (107) "Although there have been many studies of personality tests, the validity of the various tests in this field has not been established." Thus the precautionary statement earlier (100), "Before introducing a new test for measuring personality, a school will do well to try it out experimentally with a few pupils to see whether or not it contributes to the description of the personality of the individual pupil."

Several rating scales are described with a caution about "halo effect." Much more attention is given to the "anecdotal record," a systematic collection of anecdotes "concerning some aspect of pupil behavior which seems significant to the observer." No attempt is made at a statistical evaluation of this anecdotal procedure, either as to inter-observer reliability or as to the prognostic value of various types of behavior samples.

In a separate chapter, a large number of uses to which objective test results may be put are described. A discussion and proposed solution of the validity situation is given as follows (199):

> Research is the only means by which the prognostic nature of a test can be adequately investigated, but research techniques are often slow and expensive and not enough time has elapsed since the publication of the majority of the good tests for research to discover their relationships to fundamental long-time objectives. . . . Personnel workers in schools, however, cannot wait until research with tests has supplied answers to many of their questions, for they must meet today the problems of guiding their present students. Nor is it necessary for them to wait, for they can read a great deal of meaning into the scores on the basis of their own experience.

The reviewer believes that if some of the effort being spent on reading meaning into test results could be channeled into research on their relationships to objectives, the number of areas in which we read meaning *from* rather than *into* scores could be increased.

Indiana University. MERRILL ROFF.

YALE, J. R. *Frontier thinking in guidance.* Chicago: Science Research Associates, 1945. Pp. 160.

This book is an anthology of 24 articles in the field of guidance. All except one of the papers has appeared elsewhere and are reproduced in this volume because, in the opinion of the editor, they "are selections written fairly recently which the well-informed person should know or know about" and are "recommended reading of definite interest to the guidance person." Each of the 24 contributors (13 of whom are psychologists) is a recognized authority and has made significant studies in this field.

There are six parts to the book: *Part one, Education prepares to increase guidance*, contains a discussion of needed reforms in the secondary school system (Reeves), and presents proposals of the National Resources Planning Board (Blaisdell) for increased federal support and of the New York Regents (Wilson) for increased state support of secondary education; *Part two, Guidance comes of age*, contains articles on the genesis of modern guidance (Paterson), the growth of occupational testing (Borow), the relation between guidance and instruction (Brewer), evaluating the effectiveness of counseling (Wrenn and Darley), the importance of guidance in the elementary school (Addy) and includes a summary of the statement prepared by the ACE Committee on Problems and Plans in Education on student personnel services; *Part three, School guidance programs in operation*, discusses plans developed in Glencoe, Illinois (Kawin), Philadelphia (Boyer) and San Francisco (Schmaelzle); *Part four, Undertaking readjustment of the veteran*, presents material pertinent to the counseling of veterans in the post war period (Williamson), accrediting veterans returning to school (Jacobson) and two states—Connecticut and Michigan—veterans' programs (Gray, Fern, Horn); *Part five, Tools for the guidance worker*, includes articles on non-directive counseling as an effective technique (Rogers), a survey of occupational filing systems (Yale), occupational analysis activities in the War Manpower Commission (Shartle and Dvorak), a critical analysis of textbooks surveying the field of guidance and student personnel (Fredenburgh) and a minimum library for counselors (Smith); and *Part six, The counselor obtains additional training*, contains sections devoted to laboratory training for counselors (Lloyd-Jones), selecting and training teachers—in Fargo, N. D.—for individual guidance (Froehlich) and short term training programs for counselors (Jaqua).

Obviously with such a miscellaneous collection of topics it is difficult to arrive at an adequate overall evaluation. The articles were prepared quite independently of any thought of inclusion in the volume under review, so duplicate each other in minor ways and do not of course constitute a systematic treatment. Certain of the articles are critical in content whereas others are merely descriptive. They are interesting but undoubtedly many similar volumes of completely different articles could have been prepared, so one may ask, why *this* particular collection?

There is no adequate discussion of this point, nor is there any introduction or summary that brings together the points which the editor may have thought significant. It would seem that if the separate articles are worth presentation in a single volume (and the reviewer is not doubting this) they should also be worth an adequate summarization in some chapter designed to integrate the ideas involved. If the title were less pretentious and something like "A few readings in the field of guidance" the contents would be much more accurately described. This is not to deny the editor's contention, however, that some of the articles included do represent significant and critical thinking in the sense which the title implies.

LEONARD W. FERGUSON.

Field Management Division
Metropolitan Life Insurance Company
New York 10, N. Y.

DUNSMOOR, CLARENCE C. (Chairman) for New York State Counselors Association. *Practical Handbook for counselors.* Chicago: Science Research Associates, 1945. Pp. 160.

Although this handbook was prepared to assist high school counselors, it will be of interest and value to high school teachers and counselors, college and university counselors, teachers college faculty who teach courses in guidance or school administration, and to school administrators or boards who are responsible for authorizing guidance programs and selecting counselors.

An outstanding merit of the handbook is the comprehensiveness of the guidance program which it outlines. It outlines a program extending from the elementary school through follow-up studies of post-school adjustments and including educational and vocational guidance, psychological testing, personal adjustment counseling, study methods, and orientation to the school. Its major deficiency is to be found in the area of actual guidance methods. That is, the handbook presents an excellent outline of the administrative procedures, tools, and techniques of collecting data to be used in guidance but does not go very far in telling the counselor how to use the data in guiding the individual. For example, several psychological tests are listed but very little comment is made on how to interpret or use the test results; several study skills are listed but no suggestions are given as to methods by which they can be developed; and the techniques and content for a case study are outlined but no suggestions made as to how to use or interpret the data collected. It appears that the authors assumed that the counselor who uses the handbook has been trained in guidance methods and psychology but can profit by a comprehensive review of the scope of work that he should include in his guidance program. In spite of this deficiency, the handbook deserves a distinctive position in the literature regarding guidance programs in the elementary and secondary schools. If a school system organized a guidance program of the scope outlined in this handbook and secured adequately trained personnel to do the guidance work, it would have an excellent program. However, one administrative arrangement the handbook omits is for the counselor to have access to psychological or psychiatric assistance in the diagnosis and treatment of behavior problems, nor does it recommend that the counselor be trained in psychological therapy, abnormal psychology, or mental hygiene.

The handbook contains eighteen chapters, which are very brief, averaging seven pages per chapter. The chapters cover such topics as: the role of the counselor, the objectives, scope, and content of the guidance program, technique of the counseling interview, the maintenance of cumulative records, tests of abilities, achievement, interests, and personality, content and techniques of the case study, group activities as a part of the guidance program, orientation and adjustment to the school, course selection and program planning, vocational guidance, placement program, preparation for college, financing a college education, follow-up studies of post-school adjustments, the counselor's personal qualifications, the counselor's relationships with other persons and agencies, the counselor's tools, and the counselor's professional development.

Chapter 17, the *Counselor's professional tools*, contains an excellent bibliography of books, journals, and other publications dealing with youth problems, guidance techniques, specific phases of guidance, and occupational and educational information. At the end of the handbook is a bibliography of additional books on various phases of guidance and a list of agencies which establish the qualifying requirements for various occupations and professions.

WILBUR S. GREGORY.

University of Nebraska.

TRIGGS, FRANCES O. *Personnel work in schools of nursing*. Philadelphia: W. A. Saunders Co., 1945. Pp. xv+237.

The "personnel point of view" is doubtless as applicable and as efficacious in the field of nursing training as in any other area of education. In fact, a very considerable amount of work has been done in this area, as is evidenced by the bibliography of 354 titles included in the present book. The book itself is in the nature of a general survey of the field of student personnel work with particular reference to the problems of student nurses.

The general treatment of the topic is rather eclectic with a generous infusion of the "newer approach to psychotherapy" sponsored by Rogers and by the Elliotts. The consequences of this eclecticism are evident in a conflict between statistical objectivity and clinical case-work.

The book starts with a sketchy exposition of *Human Behavior and Adjustment* which runs the gamut of physiology (the autonomic nervous system and the endocrines) emotion and motivation, learning and conditioning and the psychology of adjustment—all in 23 pages. This is followed by five chapters on *The Counselling Program* (educational, personal and vocational), and three chapters on testing and personnel records. The final chapter, entitled *The Story of a Student*, sets forth a complete case history in some detail.

The treatment is rather uneven. The sections on counselling problems are written in a style that is clear and straightforward, if somewhat too cursory. In handling the more technical topics of testing and statistical analysis, however, the author often descends to a "popular" level which is confusing at best.

Altogether, the book presents a sympathetic and (for the most part) readable account of the work of the counsellor in a school of nursing which should serve as a stimulus to "start the prospective professional counsellor . . . on the road to her technical training." Its major weakness lies in its failure to provide a rational basis for the predictive process and its reliance on an uncritical amalgam of any and all information arrived at through the somewhat mystical medium of the counselling interview.

Personnel Research Section, AGO
New York, N. Y.

E. DONALD SISSON.

Vol. 43, No. 3 May, 1946

Psychological Bulletin

LANGUAGE AND PSYCHOLINGUISTICS: A REVIEW

N. H. PRONKO
Indiana University

Because of the extreme pervasiveness of language, the literature in this field is not readily amenable to systematic classification. It is only necessary to point out that language phenomena have furnished data for such diversified approaches as are represented by the anthropologist, sociologist, philosopher, educator, historian, grammarian, philologist, linguist, physicist, physiologist, and psychologist (theoretical and clinical).

With such an ubiquitous attack directed upon the manifold aspects of linguistic phenomena, and with investigators motivated by highly specific interests, one may well expect trespassing of disciplinary boundaries and possible discrepancies in their statements about language; and, while the interest of the present paper is primarily in the behavioral phases of language events, it is nevertheless considered important to specify briefly the investigative techniques employed in certain parallel fields of inquiry in the belief that such a procedure of comparison and contrast may bring the essential psychological features of linguistic happenings into sharp focus.

I. NON-PSYCHOLOGICAL STUDY OF LANGUAGE

Anthropological Interest in Language. The anthropologists' study of language is as broad as his investigation of other phases of group life, for example the presence or absence of certain skills, types of clothing, tools, ceremonies, et cetera. He may inquire into the problems of the distribution of linguistic families and the interrelationships of race, language and culture (cf. e.g., Sapir **152**, pp. 142–154); or, as Kroeber (**106**, pp. 112 ff.) has done, he may observe the relative size of vocabularies of various anthropological groups, the characteristics of the sound combinations that they employ or the diffusion and rapidity of linguistic change. On the other hand, Malinowski's (**120**, pp. 296 ff.) field studies may be said to emphasize the specificity of verbal patterning and the impossibility of translation grammar in the comparison of linguistic responses among members of different cultural groups.

Henry (76) has shown the futility of inferring about the adequacy or inadequacy of a language as judged by its formal categories, illustrating the point with an analysis of linguistic expression among the Kaingang Indians of Brazil. By means of changes in pitch and force of articulation, through modification of sound patterning by fine movements of the pharynx, and by means of gross bodily and facial changes, the Kaingang achieves great richness of expression, though lacking in formal devices.

A final but, perhaps, more typically representative study of anthropological interest in language is the recent work of Harrington (72) on the American Indian sign languages. Comparison was made between this language and Chinese ideographic writing and its prevalence was traced to the area over which the buffalo persisted longest.

Sociological Study of Language. The sociologist's pursuit of language is brought out clearly in a review of recent Soviet studies in linguistics by Schlauch (153) who relates language research in the Soviet Union to the social divisions and struggles of those who speak them and linguistic change to political and social change, while a desire to learn about earlier social conditions motivates interest in linguistic paleontology and semantics through analysis of archaisms.

Sociological influences on the patterning of language are reflected in Hargan's (71) study of prison language and in Reinecke's (145) paper, *"Pidgin English" in Hawaii: a local study in the sociology of language.*

The Philosophy of Language. Since the work of philosophers is not limited by the rigours of field and laboratory investigations adhered to by the scientific disciplines, this very lack of restrictions is an inherent characteristic of the general approach to the philosophy of language and serves at once to segregate it from the sphere of activity engaged in by psychologists and other students of language.

The problems here concern such grand questions as "the relation between the structure of language and the structure of the world" (149, p. 265) or the relation of one sentence to another or the thing or things that the first sentence stands for or means. Cassirer (25) may be placed in this general tradition when he attempts to show how language makes possible an ordered world of spatial and temporal relations, and abstract concepts of substance and cause which would not exist without language. The works of Britton (15) and of Urban (179) also belong within this somewhat ill-defined framework.

An offshoot of the preceding line of work has been a long series of "meaning of meaning" books, beginning with Ogden and Richards' (134) exposition of that topic. Korzybski (105) and a succession of popularizers [e.g., Stuart Chase (28), Hayakawa (74), Hugh Walpole (182) Aldous Huxley (86) and Irving Lee (112)], have treated the problem of meaning or semantics—the relationship of words and their referents to "clear thinking."

Educational Language Problems. Foreign language study has stimulated a great deal of interest in the most economical methods of language learning. For example: is the oral method superior to the visual mode of presenting language materials? In answer to this question, Henmon (75) found that auditory presentation of digit, non-sense and sense materials was superior to visual, visual-auditory, and to visual-auditory-motor methods. Other experiments have attempted to find the optimal rate of learning, the most favorable length of, and interval between, practice periods or the most economical language unit for study. The effect of the native language upon the acquisition of another has been investigated by Garth and Smith (60), Arsenian (4), Mitchell (126), Stark (165), and Hackman (67) with somewhat contradictory conclusions regarding the benefit of a language supplementing the native tongue.

In another type of study, Grinstead (65) learned German by two different methods: (1) looking up formal lists and (2) reading for meaning and looking up only the unknown words. He found the contextual method superior to the word-list method both in immediate and in delayed recall. Briones (14) reached somewhat the same conclusions in a similar experiment in which he compared the learning of a series of isolated words with a series of phrases or sentences. The latter, as approximations to actual speech adjustments, were superior both in learning time and in retention tests. Besides its value in possible application, this finding has implications for language theory.

Still other aspects of language learning as they concern language teaching and study have been reviewed by Powers (142), Buchanan and MacPhee (16), Huse (85), Gray (64), Tharp and McDonald (170), and Fries and Traver (57).

While, as has been mentioned, experiments stimulated by educational problems may have significance for language theory, their primary emphasis is on the practical problem of language learning and not on the nature of language behavior *per se*. In this respect, language teachers share, in part, the enigmas of the arithmetic or English teacher on the best manner of presenting a course. At any rate, such studies serve to indicate the essential difference between the educator's and the psychologist's interest in language events.

Experimental Phonetics

Physical or Acoustic Phonetics. The physicist's participation in language study consists of abstracting one small portion of the linguistic field and proceeding to a physical analysis of speech sounds *qua* sounds.

The work of Scripture is illustrative of research in this line. As early as 1904, Scripture (**154**) recorded speech sound waves by means of the phonautograph, manometric flames, the kymograph and phonograph, and made harmonic and other analysis of these data. More recently, Scripture (**157**) has built a system of "atoms" and "molecules" based upon curves obtained from a simultaneous recording of sound waves and breath pressure. He considers the syllable as a "speech molecule" consisting of separate sounds as "speech atoms" but, despite an atomistic approach. Scripture (**171**) directs criticism toward phoneticians who

> believe that the physical word consists of a series of speech sounds in which the gaps are filled with transition sounds termed "glides." All registration—macrophonic, microphonic and myokinetic—show, however that speech consists of continuous movements. There are no constant portions, no blocks of speech and no glides.*

The more general physical characteristics of speech waves have been presented in a book by Harvey Fletcher (**52**), who in addition to using the usual instrumentation has perfected the oscillograph as an aid in handling certain sound engineering problems in the Bell Telephone Laboratories as well as a tool in dissecting the pitch, volume, resonance and other physical attributes of sound waves.

Another variety of the physical approach to speech has been reported by Paget (**136**) in his *Human Speech*, in which he analyzed and artificially produced voiced and unvoiced vowels by means of double resonators made of plasticene, and has applied these findings to a theory of voice production in the human during speech. He has also made a study of vowel resonance and of certain physical features of consonants.

By an oscillographic recording of speech sounds, physical interrelationships of sound patterns have been reported by Eliason and Davis (**44**). These investigators offer an experimental demonstration of the principle proposed by historical linguists that weak stress on the final syllable facilitates lengthening of the succeeding vowel and that the intervocalic consonantal explosion tends to shorten and disappear with decreasing stress on the final syllable. These results were connected up with certain historical changes in English. A final study by Herman (77) points up the possibilities of acoustic research; for it was a harmonic analysis of oscillographic records that permitted him to pick out certain wave differentia between hoarse and non-hoarse speakers.

---

* Bloomfield (**11**, p. 76) is in agreement with Scripture on this point, for he writes: "Even a short speech is continuous: it consists of an unbroken succession of movements and sound waves. No matter into how many successive parts we break up our record for purposes of minute study, an even finer analysis is always conceivable. A speech utterance is what mathematicians call a *continuum*; it can be viewed as consisting of any desired number of successive parts."

Physiological Phonetics. The preceding account of phonetics has been restricted to the study of sound waves (linguistic physics); these sound waves are viewed as products or physical objects *produced* by a speaker, but it is also possible to probe into the sound-producing *movements* of the same speaker (linguistic physiology). The latter interest belongs to the division of physiological phonetics. Experimenters in this field have delved into the anatomico-physiological factors and such voice processes as respiration, phonation, resonance and articulation, with emphasis on the *functioning* of the speech organs.

Investigations have been carried out by means of devices such as kymographs, laryngoscopes and palatograms in an attempt to determine the exact position of tongue, soft palate, larynx, mandible and lips for the production of each speech sound.

This positional viewpoint early gave rise to a widespread belief in a fixed placement of the articulatory mechanisms in speech production, so that it was rather upsetting to phoneticians to learn of a case reported by Scripture (155) in which speech was effected without a larynx. A girl who had undergone tracheotomy at the age of three and was unable to expel *air* through the larynx could nevertheless speak distinctly but in a faint unvoiced manner. Phonautograph records showed intake of air at the beginning of speech, as in the normal speaker followed by a regular series of waves comparable to the vibrations of vocal cords in normal speakers. Further analysis showed that the "voiced" quality was produced by compressing the air in the pharynx and allowing it to escape between tongue and velum, this mechanism vibrating and functioning as an artificial glottis.

In this connection, it is interesting to mention Fletcher's (52, pp. 12–13) reports of tracheotomized persons who manage to speak well with an artificial larynx, showing that differentiation of speech sounds may be effected by mouth and lip positions also.

Another type of non-laryngeal voice production has been studied by Burger (19) and Kaiser (98). These investigators describe a subject who spoke with air from the stomach vibrating the mouth of the esophagus, producing "stomach speech" in this manner that had a range of two octaves. Besides this, Schilling and Binder (166, 167) report speech resulting from pharyngeal vibrations (i.e. buccal parts *above* the larynx). A startling series of three cases has been described only recently by Goldstein (**61**); in these cases oral speech that was understandable and quite distinct as to articulatory nicety was developed after total amputation of the tongue. The author concluded that the tongue is not an indispensable organ of articulation; that other parts of the buccal cavity may replace its function; and that our knowledge about the physiology of the tongue as an organ of speech needs extensive revision.

Despite the argument of the above clinical and experimental findings against the fixed placement of speech organs, phoneticians have, until recently, taken an uncompromising stand on the assumption of a fixed tongue position for the production of vowels. In fact, the classification of vowels in the vowel triangle as accepted by the International Phonetic Association into "open" and "closed" is based on the distance of tongue bulge from the hard palate.

As the result of an elaborate series of experiments utilizing X-ray studies of tongue position, Russell (**150**) has shown the fallacy of the physiological vowel triangle. His X-ray pictures show no consistency for tongue position of any vowel from individual to individual. Indeed he adds:

> It gave the author a distinct shock to discover that well trained *singers can sing* on a loud swelling note *almost any* of these *front vowels with the tongue relatively flat in the mouth.* Most vocal teachers would confirm this statement, and as a matter of fact, the majority of them spend considerable time in training the student to "keep the tongue out of the way" by so producing them (p. 153).

Since the physiological triangle conception of vowels is fallacious, Russell (**150**, p. 155) thinks it more valid to use an acoustic scheme to represent vowel facts.

Because the foregoing discussion of phonetics is likely to give the erroneous impression that there are two distinct camps in the field of experimental phonetics, it should be noted that there appears to be a recent eclectic trend toward the utilization of both physiological and physical speech data. Such an integrated viewpoint obviously including more features of the abstracted speech event than either alone is represented by Judson and Weaver (**97**) in their *Voice Science*, but is not entirely absent in earlier work in phonetics. Perhaps, the differences among workers in this regard may be said to be only a reflection of certain emphasis and preferences.

A more dynamic viewpoint of the physiological processes involved in speech production is evidenced in the kinesiologic approach of Stetson and his students, who stress movement of articulatory organs rather than their placement. The following quotation succinctly states Stetson's (166, p. 1) position:

> The basic movement of speech is the pulse which produces the syllable. This syllable is a puff of air through the glottis made audible by the vocal cords in speaking aloud especially in the vowels, and stopped and started by the articulatory movements of the consonants. The larynx itself does not initiate the syllable nor control the process of articulation; this is quite apparent in the case of speech with an artificial larynx. The larynx makes possible the tone of speech. In just one thing is it a prime mover; it determines the pitch of the tone. The principal problems of the speech process is above and below the larynx (sic).

Stetson (**166**, p. 38) has substantiated his theoretical stand with experimental analysis. Action current records have shown the important functioning of chest and abdominal musculature in speech generally, while Stetson and Fuller (**167**) further demonstrated, by means of kymograph records that diphthong formation is a function of the rate of articulatory movements. Carrying the work along further, Hudgins and DiCarlo (**82**) have employed similar techniques in the analysis of the change in abutting consonants as determined by the rate of operation of physiological mechanisms.

It would not be amiss to indicate at this point, studies which, however, cannot be fitted into the classificatory scheme employed here. This is the work of Luckiesh and Moss (**113**) as reported in their book, *Reading as a Visual Task.* Aside from the important technical findings the work is interesting because it demonstrates (a) how thin a connection specific investigations may have with actual language phenomena and (b) the purely physical* concern with factors in the situation under investigation. For example,

> Certain physical characteristics of the paper and ink determine the absorption of the ink. This affects the sharpness and even the actual size of details such as letters. To some extent it affects the character of the surface of the ink; that is, its dullness or glossiness. Obviously the texture of the paper also determines the fidelity of reproduction of details (**113**, p. 206).

Such physical analysis as that of Luckiesh and Moss is farthest removed from the behavioral language datum although this does not deny that printed materials may enter into linguistic behavior situations, nor does it disparage the importance of such investigation.

---

* This term is used consistently as an adjectival form of the noun, "physics," the scientific discipline and not in opposition to an alleged "spiritual."

Phonology or Practical Phonetics. The foregoing sketch of students of sound waves would exclude a large body of modern workers who are not satisfied to study speech sounds as "mere muscular movements or as disturbances in the air without regard to their use in communication" (**11**, p. 76). Their complaint is that laboratory phonetics does not enable us to connect speech sounds with meanings and unless we do resort to our everyday knowledge to tell us whether speech forms are the same or different, we cannot recognize the distinctive features of utterances on a purely phonetic level of observation. In consequence, they select as their unit of study, the phoneme, "a minimum unit of distinctive sound feature" (**11**, p. 79), which when synthesized into larger units yields the phonetic forms or structure of a language.

Workers in phonemics have devised a phonetic alphabet which consists of a distinct symbol for each phoneme of a given language. The investigator "hears out" the distinctive features in the speech of his informant and secures a transcription for further treatment. Analysis is restricted to transcriptions although phonograph records are also utilized. Techniques in this field have hardly been settled upon and disagreement among workers is common.

McDavid (**117**) urges the adoption of field recording in phonetic notation especially for dialect study, since microphones are likely to cause the informant some feeling of constraint. Besides phonograph records must be transcribed impressionistically in phonetic notation, and these transcriptions arranged and classified at additional labor and expense before they can be analyzed. A final argument against their use is the inferior recording of speech at the higher frequencies, in consequence of which McDavid urges the use of "the human ear" for both phonetic and phonemic notation.

Leaving the problem of method in this field, we progress to an examination of some current investigations in phonemics. Sebeok (**158**), for example, has constructed a phonemic system of Santali. Note that this investigator did not observe humans talking. His data were secured *in toto* from a grammar and a book of folk tales, an analysis of which permitted the determination, among other things, that "any vowel or diphthong may begin a word" and "that consonant sequences will occur within the word but never within the syllable" (**158**, p. 67).

Sebeok's study serves as an excellent illustration of the divergent interests between students of linguistic *behavior* and those of *phonemes*. The latter, handle, arrange and systematize textual materials; the former describe the action of an organism typically talking *to someone about something*. The same characterization applies to Haas' (**66**) study, in which she analyzed types of reduplication in Thai derived from sounds which they were meant to imitate (e.g., *tum tum* for the sound of a drum, et cetera). Hockett (79) has recently offered a classification of English verb forms and verb phrases, and Trager (176) a historical phonology of the Tiwa languages. A slightly philological study is that of Hall (70) who ascribes the origin of the Romance words for "flute" to cultural borrowings rather than onomatopoetic origin.

An elaborate analysis of Hidatsa texts collected by Lowie in 1911 has been recently analyzed by Harris and Voegelin (73), and Voegelin (180) has painstakingly worked out a classification of Shawnee stems.

All in all, such investigations as students of phonology have brought forth involve them in textual or transcriptional analysis. Kantor (101, pp. 46 ff.) has termed these abstracted materials "thing language." This is not to criticize interest in such specialized portions of the linguistic event but only to distinguish again the behavioral datum from other aspects of language phenomena.

II. PSYCHOLOGICAL STUDIES IN LANGUAGE

Experiments Utilizing Language-Derived Materials as Stimuli

Only brief mention need be made of a long line of experiments that began when investigators started to introduce verbal materials as stimuli. Cattell's (26) thorough study of reaction time was a pioneer attempt to use words as stimuli. Among other results, Cattell showed that reaction time to words was very little longer than that to single letters and that speed of reaction is also a function of familiarity of the printed material.

It was at about this time that verbal materials were also introduced into a succession of experiments that were motivated by a search for new "conscious elements" other than "sensations" and "feelings." This merry hunt reached its climax at about the turn of the century with the work of Buhler, Watt, Ach, Titchener (194) and others. "Imageless thought" experiments followed in profusion.

As representative of the latter, Bagley's (5) study in the apperception of the spoken sentence is cited, which also reviews earlier work along these lines. In pursuance of similar aims, Jacobson (89) asked his subjects to fixate a word for one minute and to utter it with quick repetition in an effort to get at its "meaning," which in some cases was identified with "images" evoked by the experiment. In other cases "meaning" did not come to the subject despite a wealth of "images." All in all, "meaning" was found to be an evanescent "phenomenon."

Chapin (**27**) required students to introspect their "imagery" as she read Titchener's description of *his* (Titchener's) visual imagery of meaning as "the blue grey tip of a kind of scoop which has a bit of yellow above it . . . which is just digging into a dark mass of what appears to be plastic material." Results showed wholly visual imagery present in 50 per cent of the group, wholly kinesthetic in 36.7 per cent, and visuo-kinesthetic in 13.3 per cent.

Moore (**128**) investigated the temporal relations of "meaning" and "imagery" and found that "meaning" came before the corresponding name of an object. He also found that when his subjects were asked to react now to "imagery," now to "meaning," their reaction times were much longer in the latter case.

Weld (**190**) criticized Moore for failing to distinguish "meaning" from "process" and found various types of meaning in the reports of his subjects. McDonough (**118**, p. 426), in a study stemming directly from Weld's concluded:

> The meaning is as truly a structure as the image, nor is the image the meaning. If that were so, the image would always constitute the meaning, a result which does not follow from laboratory experiments.

Sufficient account has been given of these early experiments to indicate that the general framework within which investigators worked, their purpose and their orientation were a reflection of the fashion of the times and that, as a result, such work places itself beyond the pale of language investigation defined in the widest possible sense, at the same time, they point to the incidental nature of the verbal stimuli employed.

Statistical Studies in Language

A group of studies which are, perhaps, farther removed from the language responses of individuals than the preceding are placed together because they share the same technique—namely, statistical analysis.

Using such a procedure, Zipf (**201**) has formulated a systematic approach to the observation, measurement and formulation into tentative laws of the underlying forces which "impel and direct" linguistic expression. For example, statistical study shows that the most common English word appears about once in every ten words; the next most common, once in every twenty words, et cetera, English being a mathematically harmonic language over nearly its whole extent. Similar tabulation shows that the greater the frequency of occurrence of a word in a language, the shorter the word. According to Zipf this great orderliness is an expression of an underlying force which maintains an equilibrium between variety and frequency of word usage.

Thorndike (**173**, p. 405) has questioned Zipf's conclusions as evidenced in the following quotation:

> I should think it would be in some measure a statistical artifact, adding things which lost much of their instructiveness by being combined. The relation between *a* and *b* presumably varies not only with the size of the sample, but also with the natures of the persons who speak and write, their knowledge of English, their purposes in speaking and writing, their past experiences their access to the speech and writing of others, and whatever else makes them and the language what they and it are.

Thorndike appears to object to the treatment of language responses that swamps out the manner in which an individual builds up and employs linguistic responses. Nevertheless, Zipf's work has stimulated other workers to similar research.

Using a statistical method and assuming that the relative strengths of associations in the vocabulary will be shown by the relative frequencies of associated responses, Skinner (**162**) analyzed the data reported by Kent and Rosanoff, and found that the mean frequency of the word most likely to be evoked by a given stimulus word was 258, that of the next most frequent, 121, third, 77, et cetera. The resulting curve for the 100 responses most likely to occur is approximately linear.

In another type of study, Skinner (**163**) tabulated the frequency of alliteration in 100 Shakespeare sonnets on the assumption that if their distribution were random it would argue against any "alliterative process." He found that, despite an apparent richness of alliteration, there was no significant evidence for such a process and concluded that

> so far as this aspect of poetry is concerned Shakespeare might as well have drawn his words out of a hat. The thematic or semantic forces which are responsible for the emission of speech apparently function independently of this particular formal property (p. 191).

In a study related to the harmonic series law of word frequency distribution, Carroll (**21**) derived a curve to show the relation between the total number of words in a sample and the number of different words which may be used to indicate richness or variety of vocabulary. More recently Carroll (**23**) has developed a morpheme recognition test in order to determine the relationship between knowledge of English roots and affixes to vocabulary and Latin study and has found that the study of Latin influences morpheme knowledge but not vocabulary.

Richardine and Wilson (**146**) have made a statistical analysis of diaries kept by school children and report a total of 3,407 words, plus

280 numerals and 39 letter symbols. There were 603 different words used, 59 per cent of the words appearing only once.

Mosier's (**130**) psychometric study of meaning was conducted by presenting a list of 300 adjectives selected from Thorndike's word list to 150 college students for judgment as to degree of favorableness or unfavorableness expressed. Mosier concluded that meaning is composed of two constituents, one which is fixed for all subjects and a component which varies from individual to individual.

Yule (**200**) investigated sentence length as a statistical characterisic in writing prose applying it to two disputed cases of authorship. His results were completely consonant with the view that Thomas à Kempis was and John Charlier de Gerson (a one-time contender for the authorship) was not the author of the *Imitatio* (p. 377).

Williams (**192**) made the same sort of analysis of three samples of 600 sentences from the books of H. G. Wells, G. B. Shaw and G. K. Chesterton. The findings of both investigators show that stylistic traits in written verbal responses follow the same principle of individual differences applicable to all forms of behavior.

Boder (**12**) selected 388 specimens from American writings in the area of plays, law, fiction and science and found that a ratio designated as an adjective-verb quotient showed that the number of adjectives per 100 verbs varied with the different types of material, being highest for scientific work and decreasing respectively for legal and dramatic materials.

A final study by Johnson and Reynolds (**95**) made a factor analysis of 113 college students on 10 tests covering a wide range of verbal ability, which, according to the authors, shows evidence for the existence of two hypothetical "factors," F (flow of responses) and S (selection of responses).

Studies in Language Development

Vocalization. There is a perennial interest in the manner of fixation of language responses out of the infant's early random vocalizations. Early accounts have been biographical as illustrated by the following quotation from Fenton (**51**):

> There is . . . great richness and variety in the babble as a whole. Usually it contains a far wider range of separate sound than the child will have use for in later life. Among my baby's first sounds were a number of gutturals not used in English at all, among them the German, *ch*, and a harsh-throated *gh*. The German sound of *o*, and the French, *eu*, occurred frequently, and a considerable number of slurred vowel sounds. . . .

McCarthy (**115**) has presented evidence which supports Fenton's contention for the plasticity of the infants prelinguistic "mechanisms" in language functioning, and has also set forth the views of other investigators (**116**, pp. 278–315) on this question.

Experimental attacks on the problem have been made recently by Irwin and Curry (**88**) who transcribed the vowel elements appearing in the crying of infants under ten days of age and who tabulated those vowels that were used with greatest frequency. They conclude that the infant progresses toward an increasing use of back vowels. In another paper, Irwin and Chen (**87**) have reviewed the literature on the speech sound elements during the first year of life in a month-by-month analysis.

Vocabulary and Parts of Speech. Other aspects of investigation of the linguistic developments of children have centered about vocabulary and growth of parts of speech; consequently, "word count" studies have been on the increase. Thus, Jersild and Ritzman (**92**) found that an analysis of verbatim records of children's utterances during the school day showed that the amount of talking and the size of the vocabulary both increased with age. Brigance (**13**) has reported that a comparison of parts of speech shows that as the percentage of nouns declines, the percentage of verbs increases, and has explained these results as indicating a growth of ability to express ideas in complete sentences. Further study of the developmental changes in the distribution of parts of speech by Davis (**38**) disclosed that the percentage distribution of parts of speech, over the age range from 3½ to 9½, changed very little though use of nouns and regular verbs increased with age, while use of pronouns, conjunctions, and prepositions decreased.

In a related study, Carroll (**22**), calculated the number of determining and numerating adjectives occurring in 50 responses made by 87 children of 2.5, 3.5 and 4.5 years of age in both free and controlled situations and discovered more determiners, numeratives and corresponding pronouns with age increase but the proportion of such words to the total number of words remained fairly constant. However, the average number of words used by each child showed, as with previous investigators, a significant acceleration with age increase.

Davis (**36**) has similarly found that the number of one-word remarks decreased somewhat with age, that long sentences tend to be highly complex; and that a large percentage of adverbial clauses are used and a smaller percentage of noun clauses. The children of his study also were found to use long sentences when actually referring to situations confronting them or to topics closely related to the situation. Davis (**35**) has also counted the number of proper names used by 436 children, aged 5.5, 6.5 and 9.5 in 21,800 remarks and found that the numerical importance of proper names was slight, constituting less than .5 per cent of all words used.

Language and Sex Differences. Certain investigators have attempted to find sex differences in amount, rate or patterning of linguistic responses of children. Representative of these studies is the work of Olson and Koetzle (**135**) in which combined data for two groups of nursery school and kindergarten children showed a verbal output of 16.5 words per minute (at a rate of 106 words per minute) while speaking under conditions of free play. Boys tend to speak less than girls during a given period of time but when speaking do so at a more rapid rate. Young (**198**) also confirms the verbal superiority of girls over boys.

Language Development and Environmental Factors. Pertinent here is the investigation of the language growth of children in two types of rural environments by Worbois (**196**), who found that subjects selected from a group attending a consolidated rural school showed superior verbal ability when matched with another group from a one-room school in the same area. Word recognition, word fluency and basic language skills test scores were higher in each case in the consolidated school group of subjects.

Onliness appears to be effective in inducing variety of word usage and increasing sentence length according to a study by Davis (**37**). Ninety-seven only children were found to be superior when compared with 166 twins and 173 non-only singletons. This linguistic superiority of only children was found to hold for both boys and girls.

Gansl (**58**) administered vocabulary lists to New York City public school children and found that in grades 8–14 boys were superior to girls and that there was a fairly constant increase in the vocabulary development of both sexes from the lower to the upper grades.

As a type of environmental control, Strayer (**169**) has observed the relative efficacy of early and deferred vocabulary training studied by the method of co-twin control. He concluded that

> although Twin C had reached a higher level of accomplishment in the field of language at the end of 28 days of training than Twin T had attained after 35 days of training, when she was five weeks younger, Twin C had not then caught up to Twin T's accomplishment at the close of the experiment. Follow-up contact with the twins over a period of three months after the close of the experiment, however, showed that the differences were disappearing if not entirely gone (p. 316).

Further studies in this field have been reviewed by Anderson (**3**).

Language and "Concepts"

A whole block of experimental studies may be segregated from the preceding on the basis of their subordination of the linguistic response to some other behavior, the former serving as an index, or clue to a particular "concept," "class of concepts," et cetera. Stress is on the behavior "revealed" or "expressed" by the child's language.

A paradigmatic study in this field is that of Piaget (**139**) who made detailed analysis of children's conversations in order to trace out the growth of their thinking. Piaget's findings convince him that the thought of the child evolves from an egocentric and autistic type to the logical and socialized type of adulthood.

By means of a similar method, Kuo (**110**) has recorded spontaneous language from four Chinese children three to five years old, comparing these data with verbal filling-in-blank tests given to sixteen school-children seven to nine years old, and has reported that egocentric speech, which occupied from 10 to 20 per cent of children's conversation at the earlier stage decreased with age.

In a study of the use of pronouns by young children as a clue to self-awareness, Goodenough (**62**) has discovered that pronouns of the first person singular are used more often during play with other children than when the child is alone with an adult, while pronouns of the third person singular tend to drop out with age increase.

In similar vein, Young (**199**) has analyzed language records of 72 pre-school children for use of pronouns and, from the decrease of possessive self-pronouns and increase of possessive pronouns referring to others, concludes that there is a development of "social consciousness" with age.

A discordant note has been injected into the preceding harmonious theme by a recent study of Janus (**90**, p. 56) which indicated that

> 1. At all ages from 18 to 66 months, considerably more language is devoted by children to social manipulation than to egocentric remarks. . . .
>
> 2. For a variety of play situations, children exhibit much more the tendency to evoke a response in their hearers than to verbalize for its own sake.

Janus (**90**, p. 57) concludes that language responses show marked variability at all ages and that they are, to a greater degree, a function of the play situation rather than of age difference *per se*.

In regard to the function of "egocentric speech," Vygotsky and Luria (**181**) have demonstrated that egocentric or non-social speech is likely to appear when the child is confronted by a difficult situation as a help toward organizing subsequent activity. This function of speech is never simply replaced by socialized speech but continues as a distinct process along with "external and internal speech."

Working from a different approach, Adams (1) found that among other things young children made more frequent reference to (had a "concept" of) the present and to past situations. Similarly, Davis (35) has reported that older children refer to places at a distance more frequently than younger children, who more often mention local places.

Shirley (160) has analyzed the speech of pre-school children for content, arguing that, since her subjects had not been exposed to the common environment of a nursery school, any community of response must derive from a common set of needs. These have been arranged in a hierarchy in the following order: (1) a need for succorrence or security; (2) a need for independence; (3) a need for affiliation and another for recognition which tie for third place; and (4) those for nurturance, blame-avoidance and achievement which tie for fourth. Shirley's suggestion that further analysis of common responses of children made in free conversation might yield frequency tables for use in diagnosis of personality problems shows the "thing" orientation toward language that has been employed in parts of speech and concept studies.

Difficulties with the classification of the child's linguistic responses into the rigid pigeon-holing of the grammarian had been pointed out as early as 1894 when Lukens (114, p. 454) wrote "To classify such child-words by the adult distinctions of the parts of speech . . . is of course to be misled by very superficial considerations."

Apropos this general problem, in the same year Dewey (39, p. 64) said: "The psychological classification is to class the word according to what it means to a child, not to the adult with his grammatical forms all differentiated."

Shirley's (160) study referred to above also shows that "lookit," "see," and "watch what I do" may be used interchangeably to attract the attention of others. The same holds for the question, "What's that?" which is often rhetorical and is asked for the sake of gaining attention since the child often answers it himself immediately. These are some of the dangers of confusing psychological grammar with the grammarian's grammar.

A similar stricture applies to attempts at correlating "concepts" with words used by the child. In this regard, La Brant (85, p. 393) warns that "even so small a vocabulary as that possessed by the pre-school child offers difficulties. . . . The same word or spoken symbol may serve several purposes. *Orange* is both a fruit and a color; one word to this child, two to that. *Good* has a certain meaning in *Be a good boy*, and a different one in *This apple is good.*"

The Clinical Psychology of Speech

Because the literature in this field is so voluminous, embracing a wide variety of anatomical, physiological and behavioral pathologies; and because workers here are involved only in practical considerations aimed at relieving or curing those handicapped in transmissive or receptive speech, often without any interest or need in understanding the nature of language behavior, we need not further regard clinical studies in speech.

Parenthetically, it may be mentioned that investigations of stuttering have been reviewed by Travis (**177**), Solomon (**164**), Johnson et al, (**96**), Hahn (**69**) and more recently by Hill (**78**). Weisenburg and McBride (**187**) offer both a review of the literature and experimental investigations of aphasia, and Montagu (**127**) and Whitaker (**191**) have summarized work in the field of lip reading.

Experiments in "Phonetic Symbolism"

Do the sounds of words naturally symbolize aspects of objects? Attempts to secure an answer to this question have been experimentally investigated beginning with Sapir's (**151**) study in 1929.

Sapir investigated the relation between sound and size and their corresponding word symbols by reading lists of nonsense words, differing only in one vowel, to a group of subjects who were to differentiate between the pairs on the basis of size. For example, "mal" and "mil" were arbitrarily given the meaning, "table," the subject being required to decide which of the two symbolized the large, and which, the small table. In general, Sapir found that the open vowels referred to the larger of two objects and closed vowels to the smaller in conformance with his belief that certain sounds have an "expressive symbolism" founded upon an "unconscious or intuitive logic which is not necessarily based on experience with the stimuli in their normal, functional aspect" (p. 239).

Newman (**133**) extended Sapir's study and included darkness and brightness, in addition to magnitude, reporting their differentiation through a phonetic symbolization. This was achieved by having the subjects select the larger of a pair of words differing only in their vowels (*e. g.*, glupa-glopa).

Bentley and Varon (**7**) conducted a check experiment on Sapir's and Newman's studies but with nine categories of object aspects; these included angularity, foolishness, endurance, liquidity, sentimental attachment, motion, noisiness, solidity and strength. They conclude that there is insufficient evidence for any *symbolic* values inhering in vowel or consonantal sounds as such and are of the opinion that Newman's comparisons were unwittingly made by his subjects on the direct basis of differences associated with the *sounds* themselves. Support for such a view is furnished by experimental evidence which showed that when subjects were asked to compare the magnitude of two simple sounds such as given by two tuning forks, the *lower* tone was reported "larger" than the higher in a ratio of about 7:1. With noises the *higher* noise was reported "smaller" than the *lower* in a ratio of approximately 20:1. As a final point, Bentley and Varon suggest that conceivable "magnitude" may be directly represented to the immediate visual perception of an articulating organism, the open mouth, the stretched jaw, the tensed lip, et cetera suggesting certain *associations* of magnitude. This factor may have operated in Newman's experiment but was ruled out in the Bentley and Varon investigation by screening the experimenter from the subject. These results suggest that here and there in a language certain words may have sounds with appropriate or possible *associative* values when the subject's attention is called to such associations, but to ascribe such symbolizations to the words themselves as reflections of "an unconscious or intuitive logic" is going beyond the facts, to say the least.

Another study in phonetic symbolism was performed by Fox (**56**) who presented a series of nonsense figures each accompanied by a list of six nonsense syllables, from which the subjects were required to select the most suitable name for the different figures. There was some agreement on the names chosen, especially for the sounds i, z, k, which were selected for sharp or angular figures and m, l, u and b for smooth or voluminous ones. The authors do not appear to relate these results to associative factors but do offer such a general explanation as indicated in the following statement quoted from their study: "With general instructions, O's tend to report as reasons for their choices associations involving their knowledge of languages and of familiar physical objects (**56**, p. 578)."

Panconelli's (**138**) results are more in line with those of the Bentley and Varon study. This author subjected natives of Hamburg to the noises of sawing and pounding. The subjects reproduced the stimuli with widely varying symbols that corresponded more to the "hearer's ideational world" than to the actual auditory stimulus. Consequently, Panconelli believes that he has refuted Sullivan's theory which claims a close relationship between the sounds of natural phenomena (such as thunder) and their phonetic symbolization.

It may well be that the reason why these findings contradict those of Sapir and Newman lies in the methods employed since Panconelli's subjects had no associative aids from nonsense syllables that inevitably bear some resemblance to actual words. Another criticism against claims of the appropriateness of verbal symbols to the object symbolized has been raised by Kantor (**101**, p. 171) who points out the prejudiced selection of cases that completely ignores exceptions to the alleged rule. A final point pertaining to such studies in phonetic symbolism may be put as follows: How are such experiments related to the language actions of organisms? Do they throw light on such behavior?

Experiments in Language Employing the Association Technique

Repeating an earlier experiment of Thumb and Marbe, Esper (**46**) used association reaction time and found that the more common a given response, the shorter its reaction time and that verbal stimuli tend to be associated with word responses belonging to the same grammatical category (e.g., adjectives are associated with adjectives, numerals with numerals, et cetera).

Another experiment by Esper (**47**) was designed to study associative interferences and the development of linguistic categories in artificial language materials. Four shapes and four colors were used in 16 different combinations, each of which had a nonsense name assigned to it consisting of two syllables correlated throughout with the two factors of shape and color. The order of the linguistic "elements" was color-shape, conforming to the English grammatical construction, adjective-noun (e.g., blue triangle). Results showed that the names of figures confused with one another during learning tended to interfere with each other in articulation. Greatest associative interference obtained in the case of linguistic material deviating from customary modes of syllable division. It was also found that names of figures of a given category tended to be confused and assimilated to the name of a particular figure of that category.

Adopting Esper's technique, Wolfle (**193**) analyzed the errors made in naming nonsense figures, such names consisting of size-name and form-name elements, the size-name syllable coming first. Analogic changes made both during learning and in the retention tests were of two kinds: those in which a subject viewing figure A would give it the name belonging to B, and those in which he would respond with a mixture of names belonging to A and B. Associations followed the categories of size more than of form, and 55 per cent of all errors were due to interference on the part of some other figure of the same shape as the one on which the error was made.

Since it was not clear if the first syllable in the dissyllabic names used was more unstable because of its position or if such instability was inherent in size names, Wolfle (**193**) repeated the above experiment reversing the order of the syllables and concluded that "the position of a syllable is of much greater importance than its meaning in determining its stability or instability" (p. 315).

A subsequent experiment by Esper (**48**) showed that instability of verbal responses occurred when stimulus objects, because of common features, set up an associative interference so that the sound or articulation corresponding to one evokes the symbol for the other one, such associative interference being indicated by "simultaneous or successive activation by one stimulus object of two or more response mechanisms, and where the tendency is strong, by lengthened reaction time in the verbal response to the object (p. 378)."

Skinner and others have perfected several ingenious devices for studying phrase or sentence associations rather than words as in Esper's or Wolfle's studies. Skinner's (**161**) verbal summator is a phonograph record which presents faint and indifferent sound sequences to the subject who is urged to hear them as words, phrases or sentences. Continued reaction usually elicits some associated verbal response from the subject's repertoire of verbal behavior.

Estes (**50**) has developed a visual form of the verbal summator which exposes skeletonized words, phrases or sentences presented tachistoscopically for 1/10 second.

The fact that the subject is urged to report the "first thing that flashes through his mind" places these experiments in the field of suggestion although Skinner and Estes explain the subject's behavior in terms of a response to the skeletal material which somehow "summates" with a strong "latent response," the response serving as an index of the "strength" of such latent response. The associative value of Skinner's technique has recently been adapted to the diagnosis of abnormal behavior by Trussell (**178**) after the manner of the Rorschach test.

"An experimental study of the effect of language on the reproduction of visually perceived form" by Carmichael, Hogan and Walter (**20**) bears certain resemblances to the techniques utilized by Skinner and his colleagues, in that suggestion and association is involved in both. In this experiment a set of 12 relatively ambiguous figures was presented to subjects. However, one list of names was given to the figures just before presentation to one group while another list was applied to the same figures prior to exposure to another group. Later reproduction of the figures by the subjects was generally in the direction of the names applied to them. According to these authors, the results "may at least tentatively be described in terms of a dynamically considered process of association" (p. 86).

Starting with the assumption that language is a form of conditioned behavior, and reinforced by the phenomenon of generalization gleaned from experiments dealing with verbal and other conditioning stimuli, Cofer and Foley (29) recently lay the theoretical groundwork for a series of experiments involving the word association technique.

As a follow-up study, (53) they attempted to determine whether or not generalization in word associations could occur along more than one homophone gradient from a given word and along synonym gradients and synonym gradients two transformations removed from the original word. Their findings showed generalization along the predicted dimensions in support of their theoretical analysis.

Subsequent work by Cofer, Janis and Rowell (30) demonstrated generalization along antonym gradients, while Foley and Mathews (55) showed a greater degree in mediated generalization via Spanish words, such "gradients" having been established for their subjects during a regular course of instruction in Spanish. The Control group with significantly lower recall scores received no such "reinforcement."

Recently Foley and MacMillan (54) showed differences in the type of verbal responses given by members of different professional groups "thus affirming the occupational conditioning of verbal reactions and suggesting the possibility of operation of mediated generalization along 'professional gradients' (p. 309.)"

Belonging to this general area of investigations dealing with verbal materials, is an experiment in "Some Complications of Associative Processes" by Thorndike (175). Thorndike presented a list of 112 paired associates, many members of the first pair of which were rhymed words. His findings in the recall test showed that a mental set had been created which tended to elicit many unheard of, and meaningless words (this set apparently created by the meaninglessness of the connection between the paired associates). Many words similar in meaning to the *second* members of the pairs and many rhyme responses were also the result of "mental set" created by conditions present during the reading of word-stimulus pairs.

That words may function as substitute stimuli the same as other objects is apparent from another study of Thorndike's (172) in which subjects were asked to rate words for pleasantness or unpleasantness of the sound alone without regard to their meaning, such responses showing influence of past associations. This was confirmed by further use of nonsense words which were rated pleasant or unpleasant on the basis of their *resemblance* to meaningful words that had one or the other type of associative value. The differences commonly assumed for the superiority of vowels and liquids and the inferiority of gutturals and aspirates were disproved by artificial words composed of those sounds. The controlling factor was association.

Gesture

> Gestures as a species of human behavior have been sadly neglected. The greatest of psychologists, Wundt and Freud, have not exactly neglected them, but have done little more than mention them. Until recently no one tried to study them experimentally, and even now there are more problems connected with them than any psychologist could hope to solve in a life-time.
>
> —M. H. Krout (**109**)

One need not agree with Krout *in toto* to acknowledge the indifference of psychologists to the functioning of gestures in linguistic responses. While Kantor (**101**, p. 160) has indicated the equivalence of gestural to buccal action in referor language behavior and has criticized linguists for failing to incorporate them into language study, yet observational or experimental studies of gestural language are practically non-existent.

Krout (**108**), for one, has made an analysis of the social and psychological significance of gestures differentiating acultural gestures such as pre-natal, congenital, animal and infantile movements, from conventional, pseudo-conventional and autistic gestures. He (**107**) has also experimentally investigated autistic gestures, developing a method for observing them and testing the consistency with which they reappear. At the same time, Krout has demonstrated that autistic gestures originate in conflict situations.

Critchley (**32**) has done no experimental work but has called attention to gestural modes of communication which may be used independently of, or in connection with, oral speech. He is of the opinion that besides symbolic gestures of a conventional sort that have been culturally established, there are also instinctive gestures comprehensible to all in utter disregard of international boundaries.

Paget (**136**) has estimated that it is possible to make 700,000 differentiable gestures by combining postures and movements of upper arm, forearm, wrist and fingers, which would make the hands 20,000 times more versatile than the mouth. In view of the possibilities, Paget has experimented with the development of an easily learned universal gesture language that is based upon Ogden's Basic English.

Efron (**42**) has made a pioneer study of the frequency and kinds of gestures among Jews and southern Italians. He has compared members of these two groups who were each more or less culturally isolated with those who have been assimilated, and has pointed out differences in the style and frequency of gesturing and its elimination in the assimilated members of both groups. Since Efron's work is anthropological in scope, there is little concern with the psychological nature of gesture; rather, his interest is in the differences of gestural patterning in different groups.

All in all, investigations of the gestural aspects of language responses have hardly had a beginning; nevertheless, it would seem that this field has possibilities of development once proper techniques have been worked out.

Studies in the Perception of Language

The factors pertaining to the perception of linguistic stimuli have been investigated by several students with rather interesting results and implications. For example, Howells (**80**) and Howells and Schooland (**81**) have worked out a test consisting of standard speech stimuli recorded on phonograph records along with other conflicting sounds, played loudly to 100 subjects, the number of words perceived by each subject constituting his score. Speech perception scores showed negligible correlation with two reliable tests of auditory acuity, indicating that the acoustic factor in the perception of speech has been greatly overestimated. These investigators conclude that: "The common assumption that ability to hear speech is analyzable into elementary factors, or acuity thresholds, is unwarranted on the basis of available evidence (**81**, p. 345)."

Preston (**143**) analyzed the speed of word perception and its relation to reading ability, and discovered a significant difference between the perception times for familiar and unfamiliar words. He also found that perception speeds showed a relation to the subject's speed of reading and to his vocabulary knowledge.

A final experiment by Cotton (**31**) concludes the type of investigation carried on under this category. Assuming a preponderant factor of lip reading in the ordinary perception of speech by normal individuals. Cotton used a small, sound-proofed room with double glass windows and made quantitative measurement of this visuo-perceptual component of speech reception. The speaker inside the room could be made visible by varying the inside illumination; at the same time, his speech could be distorted. In general, the results showed that when speech distortion occurred while the speaker was not visible, his speech was practically unintelligible to the subjects; but proper comprehension resulted from the same amount of distortion when the speaker was simply made visible to the audience. While not, by any means, complete this brief survey of perceptual problems investigated shows some of the factors related to language behavior and serves to place the latter in its proper perspective.

A Study in Language Patterning

Lately Thorndike (174) has investigated some of the functional features of language in an attempt to find the changes in meaning that a word undergoes when it is changed grammatically as from adjective to verb (e.g. to beautiful). He found the comprehension of such radical modification of speech patterning to be heavily dependent upon the immediate situation. In applying his findings to historical changes of words toward abbreviated forms, Thorndike believes that economy of effort rather than desire for variety motivate the truncation of words with frequent use, and thinks that desire for word variety is a cultivated want. In interpreting the avoidance of too frequent repetition of words by literary people, Thorndike suggests that "the explanation probably lies in a rather fundamental tendency of the neurones to be repugnant to action after a certain number of actions within a given time (p. 37)."

A Study in Language Behavior

Starting from a somewhat different orientation, Bucklew (17) has recently designed a number of experiments to study speech reception. The first experiment was meant to determine the speed with which language serves to refer individuals to the things spoken about. A series of 10 slides was presented, (a) simultaneous with the beginning of speech to one group of subjects, and (b) at the end of speech for the other group. The speech, presented phonographically, referred to various aspects of the illuminated slides confronting the subject. Time of response was measured by means of a chronograph from the beginning of speech until the subject pressed a true or false key indicating the truth or falsity of the statements pertaining to the slides. Results showed an average increase of .8 second from concurrent to serial presentation; consequently, Bucklew concluded that language functioned more efficiently when the slide was presented at the beginning of speech and related these results to an Interbehavioral theory of language.

A second experiment attempted to find whether speech distortion disrupted a language response more when the stimulus object referred to was present or absent. Ten sentences, a normal and a distorted version of each, were phonographically presented to 66 subjects. When the stimulus object referred to was absent, the percentage of errors increased to three and one half times that for the concrete situation (stimulus object present). Introspective reports showed that about 4 out of every 5 subjects reported the distortion more disturbing in the former situation.

A check was made upon Experiment II by distorting the *same* words in both concrete and remote situations in an identical fashion; it was found that there was no significant rise in RT in the concrete situation but an average increase of .36 seconds was found for the remote situation, while per cent errors rose to 13 per cent. Introspective reports also showed that about twice as many subjects judged the remote distortion more disturbing than the concrete. Bucklew concluded that speech distortion is less likely to interfere in speech reception where the referent is before the subject.*

III. THEORY IN PSYCHOLINGUISTICS

The Need for Language Construction

The preceding account of studies in language proves that there has been no dearth of research in this area. Apparently too, students of language have been about as ingenious in devising apparatus and techniques and in designing experiments as investigators in other fields. Yet, despite a wealth of interest and of work, there seems to be no corresponding increase in our understanding of linguistic responses. Descriptions of such common, ordinary every-day behavior as occurs in the case of two human organisms conversing (behavior, incidentally, which infiltrates all other forms of behavior) is conspicuously absent in our most up-to-date text books in psychology (**41, 91, 124, 194**) or where developed in scanty fashion (**129, 131, 148**), they follow traditional prescriptions known equally well to the man in the street through his cultural heritage.

---

* This survey of behavioral investigations in language is, by no means, exhaustive. The attempt has been to give a sketch of the range of problems included and modes of attack employed. However, further references are available in the work of Esper (**46, 49**) Adams and Powers (**2**), Kantor (**99–104**), and McGranahan (**119**) and Metcalf (**125**).

Such a state of affairs may well be the reflection of a lack of proper orientation toward linguistic behavior. It is quite likely that what theory† has been developed has been found disserviceable in unifying linguistic research and in pointing the way to further, fruitful investigation. And, again, where psychologists have either ignored or shunned theory, or failed to make explicit their assumptions about language, they have been prevented from attaining a first, crude approximation at the isolation of language responses and the segregation of them from other classes of behavior. As an example of such uncritical procedure, there is the case of a recent writer (**144**, Chap. 6) whose implicit (and unwitting?) attachment to an "Expressionistic" theory of language has caused him to include such disparate behaviors as block-building, drawing, and mathematics within the language category.

Further testimony of the psychologist's incompetence in the language area has come from many quarters. Speaking with reference to the need for a psychology of language that might have usefulness in language teaching, Buchanan and MacPhee (**16**, p. 25) write as follows: "The present interest of language teachers in psychology . . . is frequently a source of embarrassment to the conscientious psychologist."

Huse (**85**) has also presented a destructive criticism from language teachers concerning (a) the best methods and materials for language learning to be deduced from the contradictory findings of experimentalists and (b) the apparent lack of understanding of the psychological processes involved in language.

Woolbert (**195**) and Eisenson (**43**) have both expressed a like need in the teaching of public speaking. This awareness of the indispensableness of some sort of psychological basis for their own procedure on the part of non-psychological students is not difficult to understand when we note that all investigators draw upon the same event, namely, humans talking to some one about something (i.e., responding to stimuli). Their individual abstractions require a psychological description at many points for filling in the completed picture. Thus, Bloomfield (**11**, p. 78) has surrendered the problem of "meaning" entirely to the psychologist while Jespersen (**94**, p. 17–18) is found involved in the considerations of such activities as speaking and hearing and other behavioral items. Another case in point is Hagboldt's (**68**, pp. 84–91) reflections on the psychological aspects of reading. Palmer (**137**) and Gray (**63**) have expressed similar dependence on psychology and, among

---

† The term, theory, is not used in the present paper in the sense of a statement in terms of some hypothetical correlate of language action, but refers to an observationally-derived description of what happens when organisms behave linguistically.

among psychologists, Esper (49) has shown a sensitivity to the need for language construction.

Although a need for linguistic orientation has been expressed on all sides, language construction has been of an inadequate sort at best, so that as a way out of this predicament, people in related fields of language study have done one of two things. They have, in some cases, taken whatever feeble theoretical framework has been available or else they have gone ahead in constructing their own behavior foundations without the psychologist's help. In either case, insofar as such workers involve themselves in behavioral consideration, they become legitimate matter for our observation. It would seem worthwhile, then, to evaluate the present status of language construction from the psychological viewpoint. To this end, we turn to a review of sample definitions of language wherever they may be found, without distinguishing between the technical psychologist's construction from that of the worker in related fields. Whatever enters the field of language construction in its behavioral aspects is a legitimate object for our examination and dissection.

Dualistic Language Theory

Mentalistic language theory may take one of two forms. Where stress is on the "psychic" *deus ex machina* that converts things into thoughts, thoughts into word images, et cetera, we have the typical expressionistic design. Where prominence is given words as symbols mediating between a mental state and a thing, it is indicative of symbolic construction.

Expressionistic Construction

Wundt. Language as the instrument for the expression of thought was first systematically developed by Wundt (197, pp. 60–61) who traces its origin to gestures involuntarily expressing strong feelings. These arouse corresponding feelings in others. The displacement of emotional by ideational "psychic" components, which become voluntary, leads to an exchange of thought; this is language. Darwinian evolutionary influences are everywhere apparent in Wund't attempts at language construction.

Bloomfield (1914). At this stage of his development, Bloomfield (8, p. vi) builds squarely upon Wundt's foundation. The following quotation clearly shows the Wundtian stamp and places him at once within the "Expressionistic" fold. In addition, it demonstrates Bloomfield's investigative procedure from such a psychological base.

True to its original form of an outcry under the most violent experiences, language is most realized under emotional stress. Some violence of experience must normally be present to call forth loud expression. If this emotional violence is the dominant cause of the utterance, we speak of exclamation (p. 70). With communicative motive, this is weakened to *declarative* utterance, and further to a *question*, which is more of a "transference to communicative use of the exclamation" (pp. 70–71).

Stinchfield. Stinchfield (**168**) *assumes* a "language" without any specification of the matter under consideration, but describes the machinery for an expressionistic theory when she details the transformational processes by which "sensations" "arouse" "images,"—clusters of the latter composing "ideas." "Words or symbols which stand for ideas are stored in the sensory areas for vision and audition, and in the motor speech area, whence the impulse goes down over the motor pathway (p. 90)."

Pillsbury and Meader. One of the rare textbooks which attempts technical analysis of language from the psychological viewpoint is by Pillsbury and Meader (**141**). This work follows the orthodox Expression doctrine when it construes language as "a means or instrument for the communication of thought, including ideas and emotions" (p. 4). The exact details by which ideas are converted into language are well worth quoting.

> . . . man thinks first and then expresses his thought in words by some sort of translation. To understand this it is necessary to know how the words present themselves in the consciousness of the individual, how they are related to ideas of another type than the verbal, how the ideas originate and how they arouse the words as images, how the movements of speech are evoked by these ideas, and finally how the listener or reader translates the words that he hears or the word that he sees into thoughts of his own.
>
> Speech has its origin in the mind of the speaker or writer and the process of communication is completed only when the word uttered or spoken arouses an idea in the listener or reader (pp. 92–93).

Pillsbury and Meader (**141**, p. 16, Fig. 1) offer a diagramatic representation of the elaborate speech processes as they have described them above. The train of events begins in the speaker or writer with thoughts or feelings and concomitant mental imagery, which are shown to innervate the muscles of the lips, tongue (or hands in the case of writing). The result is either ether vibrations or air vibrations carried to the eye or ear (respectively) and to the hearer's brain where they are *transformed* into "sensations." These last, then, give rise to an "idea" corresponding to the "idea" in the speaker that started the whole series of mystical operations.

Gardiner. Gardiner (59) has criticized Expressionistic theories of language because "expression of thought" is such a minor function of language, more common ones being "to inform somebody about something, to exhort somebody to do something, or to win sympathy from somebody in respect of something" (p. 64).

So conceived, speech is said to comprise four factors: the speaker, the listener, the thing spoken of, *and* the actual *words*. Since Gardiner has reified the speaker's action into "word-signs" (p. 76) or "clues" (p. 80) his theory may therefore also be classed under symbolic types of construction but more detailed specification shows that in a typical situation (pp. 62–80) the speaker, *having* a "thought" of some *thing* (this transformation being essential), gets a corresponding "sound image" of a particular word, then volitions to articulate it, thus "transmuting the psychic word" (Fig. 4, p. 91) into a physical disturbance which operates in like fashion for the hearer. Essentially the machinery constructed here easily fits the Expressionistic mold, and although a dual classification is possible, Gardiner is placed in this category as a matter of convenience.

Gray. Since constructional details of the Expression doctrine show little variation from author to author, it will be sufficient merely to quote Gray's (63) general approach to language which is conceived as "a physical and external manifestation of a non-physical (emotional, intellectual, spiritual) and internal state, an endeavor to represent materially what is essentially immaterial (p. 15)."*

Lest the reader think that such construction belongs to the hoary past, we again call attention to the aforementioned 1941 publication (144) to show the insidiousness with which this theory has crept into current psychological thinking. It will be recalled how block-building, drawing and mathematics were identified with language much after the manner of Croce's (33) extension of Wundt's Expression idea to art in the former's identification of the linguistic with the esthetic.

The penetrating influence of Expression doctrine is grossly apparent in the most recent text books, as the following examples from Child Psychology demonstrate. For one, Hurlock, 1942 (84, p. 157) uses the term, language, to refer to "every means of communication in which thought and feelings are symbolized so as to convey meaning, involving such widely differing forms of communication as written, spoken, sign, facial gesture, pantomine, and art (sic)." Speech is only one sub-type of language that requires articulate sounds or words to convey thoughts.

---

* Attention is directed to the date of publication of Gray's *Foundation of language*, which is A.D., 1939.

Nagge (**132**, p. 274) is also seen to cling to such out-moded theory when he states that "vocal language is the means whereby words become a method of expression of ideas and feelings." Further evidence of Expressionism is to be noted in his "animal grunt" theory of the origin of language, deriving from Wundt in his definition of a sentence which is said to be "a chain or series of words expressing a complete thought" (p. 285).

Expression theory is, therefore, by no means dead. Viewpoints in psychology as in other sciences, once established, die hard. One perpetuating factor is a pretended or avowed disinterest in language theory on the part of text book writers, an indifference which militates against a sensitivity to currents of thinking in the cultural atmosphere that consequently insinuate themselves into psychological writings.

Symbolic Theory

Jespersen. With keen discernment, Jespersen (**93**, p. 7) has realized that language has no disembodied existence—that it is not an entity in the sense of a beach or a dog but must be conceived as a function of human beings—and although he considers communications in the light of "habits," it is a dualistic psychology that forces him to treat language actions as "things," e.g.,

> A word . . . may be rightly compared with such an habitual act as taking off one's hat or raising one's fingers to one's cap: in both cases we have a certain set of muscular activities which, when seen or heard by somebody else shows him what is passing in the mind of the original agent (sic!) or what he desires to bring to the consciousness of the other man (or men) (p. 8).

Language is, therefore, no longer "action" but is reified as a series of things or symbols standing for, or evoking thought. Incidentally, one may see Expression doctrine rearing its ugly head in the near background.

Ogden and Richards. Besides having emotive and other functions, according to Ogden and Richards (**134**), language also performs the symbolic function of standing for things. But the relationship between symbol and referent is not direct; for there is an intervening factor, a "thought," which mediates between the two. Grammatical study imputes a direct connection between symbol and referent but the connection is really an indirect one around the two sides of a triangle (**134**, p. 11) with "thought" at the apex, and "symbol" and "referent" at each side of the base (the base line being shown dotted to illustrate the indirect relationship). In addition to objectionable, psychic implications inherent in such theory, there exists the confusion of genuine symbolic behavior with linguistic responses as indicated by Kantor (**102**, pp. 66, 67, 103, 104).

Sapir. For Sapir (**152**) it is more satisfying to study the more dependable outward garments of language but its psychic correlate must be taken into account at the very outset as noted in his definition of language, which is a

> . . . purely human and non-instinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols (p. 7).

How these symbols work—

> the speech element "house" is the symbol, first and foremost, not of a single perception, nor even of the notion of a particular object, but of a "concept," in other words, of a convenient capsule of thought that embraces thousands of distinct experiences and that is ready to take in thousands more (p. 12).

How "concepts" sometimes evaporate—

> Each element in the sentence defines a separate concept or conceptual relation or both combined, but the sentence as a whole has no conceptual significance whatever (p. 13).

Pieron. There is no organized presentation of language as symbolism in a book treating with other problems but because of the intimate connection between language and thought demanded by a symbol theory that has been incorporated within the work, their kinship comes out here and there. On p. 30, for example, Pieron (**140**) refers to language "as the principal form of symbolization" and on p. 48 it is "the essential instrument of 'symbolic thought'." This, at least, sketches the outline of a theory developed more fully elsewhere by this and other authors.

Buhler. For Buhler (**18**) language has three functions,—Expression, Evocation, and Representation. The *expression* function is analyzed out from a consideration of the conversion of objects into thoughts and thoughts into articulations from the viewpoint of the speaker. From the standpoint of the hearer, speech *Evokes* thoughts or ideas of the things which it symbolizes. While these two functions of language are important, its nature has not been described until one has given an account of *Representation*, which is the relation between the speaker's action here reified into conventionalized symbols and the things referred to. The two fields which are important in Buhler's elaborate development of the Representational function of language are the Demonstration Field (Zeigfeld), in which the Expressor and the Evoker of language are oriented and the Symbol Field (Symbolfeld) constituting the linguistic factors in which the symbols are imbedded. The latter includes the more formal features of grammar such as case and tense and the associated factors or "spheres" which certain words carry as when "*accident*" suggests "automobile."

Palmer (**137**). Language defined:

> Language has properly three functions: it expresses the thoughts, feelings, etc. of the speaker; it influences the behavior (in its widest sense) of the hearer (the evocative or affective function); and it symbolizes the "things" to which it refers. Since the first two functions cannot be carried out without the last, this means that symbolization, the peculiar alignment of sound and meaning, is the constitutive factor of language. Language may then be defined as the symbolic system in which vocal sounds constitute the material of the symbols (p. 1).

The resemblance to Buhler's theory is quite obvious.

Eisenson. In this recent (1940) text book on *The psychology of speech*, Eisenson (**43**, p. 3) asserts that "language is any system of recognized symbols used to produce or prevent specific responses of thought, or feelings, or actions." A symbol is further defined as standing for something else. "It implies that a connection of some sort has been made between a word and an object, or a word and an idea. . . . The word is the sign for the object or idea (p. 4)."

Summary of Symbol Expression Theories

A review of the mentalistic procedure along constructional lines shows that a characteristic feature of their work is the bifurcation of a speech act into two discrete components—an inner psychic "thought" or "idea" and an outward material correlate, the "word" or "symbol." Fundamentally there is little difference between them. Such differences as exist involve only points of emphasis and elaboration of an analysis which has succeeded in resolving language behavior into *form* and *content*. Where emphasis is on the form side which considers how thoughts get transmuted into words and words into thoughts, we have Expression theory such as Pillsbury and Meader, Gardiner, and Gray previously referred to have developed. Where Expression theory generates insecurity in dealing with the "inner phase" of language because "its inner meaning, its psychic value or intensity, varies freely with attention or the selective interest of the mind, also, needless to say, with the mind's general development (**152**, p. 13)." then one stresses words, "the outer garment of thought" (p. 14). In either case, one always has one or the other partner of the psychic bi-polar "thought-symbol." One escape from this dilemma is behavioristic construction to which attention is next directed.

Behavioristic Attempts at Language Construction

Watson. Behavioristic language construction has shown distinct progress in the direction of objectivity by extruding all psychics from the description of language action. As early as 1914 Watson (**183**, p. 328) insisted on placing language habits on a par with other bodily habits. In *Psychology from the Standpoint of a Behaviorist* (**184**, Chap. IX) language habits were anchored down to "the whole body but specifically to the neuro-muscular system in the head, neck and chest segments" (p. 310). There is little technical improvement here, however, since Watson's chief purpose was to connect implicit language habits (thinking) with explicit speech habits; but, by 1930, in his book, *Behaviorism*, (**185**, pp. 228 ff.) language habits are fitted into the conditioned response framework adopted from Pavlov. Except for this rigid machinery, Watson may have come close to an adjustmental view of language action. Under the circumstances, words are treated as substitutes for objects that become organized into a series of conditioned responses. Muscular stimuli come to serve in place of visual stimuli as illustrated in the following quotation:

> Suppose you read from your little book (your mother usually sets an auditory pattern), "Now—I—lay—me—down—to—sleep." The sight of "now" brings the saying of "now" (response 1) the sight of "I," response of saying "I" (response 2) and so on throughout the series. Soon the mere saying of "now" becomes the motor (kinaesthetic) stimulus for saying "I" (**185**, p. 235).

It is to be assumed that the hearer responds similarly through a series of responses in the same manner as the speaker advanced beyond the stage illustrated above.* With this approach, language is reduced to a mechanical series of conditioned responses; although it is questionable if such a formulation does full justice to the language action of an individual engaged in conversation. We are in sympathy with Seth and Guthrie (**159**, p. 76) who ask *a propos* this general problem:

> How is it then that the right words occur? How is it that speech keeps to the rails of common sense and intelligibility? Mere association . . . is as likely as not to take us off at a tangent into incomprehensibility. If I start out to say, *The pack of cards is on the table*, the mere activity of association is as likely to carry me from the word *pack*, to *hounds*, or *wolves* as to cards.†

---

* Watson's (**186**, Vol. 3, p. 329) recent statements on language are substantially unchanged.

† These authors show insight into the importance of the general speech situation in guiding the speech pattern of the speaker when they mention his "attention upon a particular object, the 'thing meant' " (**159**, p. 76).

Is not our everyday conversation more like the adventure of faith shown by the British orator, Charles Fox, who got himself into the middle of a sentence and trusted Providence to get him out again?

A chief complaint against Watson's theory is that it leaves more unsaid than explained. It circumvents problems by refusing to face them squarely, avoiding them merely by scantiness of statement. Robinson's (**147**, p. 86) criticism is quite pertinent because it points out how inadequate Watson's mechanistic-limited system is in dealing with behavior events; therefore, we quote him fully.

> On the ordinary view, which admits the appeal to the consciousness of the agent,* the child has learnt that the word "'means' the 'thing' . . . ." But the Behavorist has no business to know that the child "desires" or "wants" the box. The consequence is that if the Behaviorist is to be faithful to his initial abstractions and keep within the limits he has laid down, he must define his substitute response simply by the place it occupies in the general "pattern" formed by all responses, implicit and explicit. But what is there in the nature of a laryngeal response as such which can possibly label it as a substitute for the response of "the larger musculature"? The only differentiating circumstances possible are order in time and frequency of accompaniment and their insufficiency is too patent to need comment. Consequently, the notion of substitution does not seem to work; it is by no means clear what "relevancy" could mean in the theory under discussion.

In the same symposium dealing with the identity of thinking and language, the substitute nature of Watson's "response" is again discussed by Thomson (**171**, p. 65) who interprets Watson to mean by substitution a "relevant substitution." Quoting Watson's account of a child learning to say "box" instead of reaching for box, Thomson asserts:

> "Box" is not true language when it merely accompanies the thing, but only when "the child learns that the uttered word is a sufficient stimulus to cause the attendant to hand him the objects he names without actually having to execute bodily movements with respect to them" (p. 66).

A final comment by Bartlett and Smith (**6**, p. 56) sums up the discussion on the inadequacy of Watson's concept of substitution. That Watson implies something more than his postulatory set-up permits is apparent in the suggestion made by these writers that a conditioned reflex level of functioning might establish a vocal habit at best but could not contain, in addition, "later associative connection of the word when learned with the bodily habits connected with the object for which the word stands" Watson, (**184**, p. 320).

---

* Although the mentalistic system espoused by Robinson is no more satisfactory inasmuch as it treats language behavior in antinaturalistic terms, his ability to handle the speech event within his postulatory system must be admitted.

In the writer's opinion, Watson would like to eat his cake and have it, too. When he finds that his system is inadequate for handling language responses as conditioned reflexes, he goes beyond the confines defined by his postulates and squeezes in an explanatory statement legitimately belonging to viewpoints regulated by radically different postulates. While not squarely faced and developed, his acknowledged realization of the "relevant" substitute-stimulus function of words, permits classifying his rather dilute language construction within the symbolic category.

A final point is to indicate the substitute stimulus idea germinant in Watson's construction but waiting for full development by future "Behaviorists." That he did not advance his construction more elaborately may be appreciated from his serious attempt to objectify thinking by identifying it with language.

Weiss. Weiss introduces a variant feature into behavioristic language theory with his stress on the social function of speech and subordination of the anatomical involvements of language responses. The dual character of the linguistic response functioning either as a response to a stimulus or as a stimulus for a response is said to be an advantage for social cooperation and communication. Its substitute stimulus function is developed in *Linguistics and Psychology* (**188**, pp. 53 ff.) as well as in his later text. In the latter (**189**, p. 313) a word "becomes a bio-social stimulus (symbol or signal) for sensori-motor interchanges between individuals."* Again, "the language response as a substitute stimulus for actual objects or situations allows the bio-social reproduction of practically any stimulating conditions that have ever occurred at any time or place" (p. 324). In short, if one overlooks the social emphasis that Weiss gives to language, his formulation is not essentially different from that of Watson; and, while the following statement is taken from Weiss (**189**, p. 288) it may characterize both, for it demonstrates the mechanical and serial action which both ascribed to language: "When I respond to the voice of my fellow man," Weiss writes, "I react to the series of sounds that he produces."

Bloomfield—since 1925. Feigning an indifference to psychology, Bloomfield (**11**, p. 153) throws overboard the mentalistic system of Wundt which did service for him in his 1914 text (reviewed above) and proceeds alone in his analysis, although here and there in his writings one may see traces of Weiss (cf., e.g., his *Why a Linguistic Society*?) (**9**).

In his *A Set of Postulates for the Science of Language*, Bloomfield (**10**, pp. 154–157) starts from a foundation in which the existence and interaction of social groups held together by language is granted by psychology. Furthermore psychology gives this series:

---

* Mead's (**123**) influence is here apparent.

> to certain stimuli (A) a person reacts by speaking; his speech (B) in turn stimulates his hearers to certain reactions (C). By a social habit which every person acquires from his elders, A-B-C are closely correlated. Within this correlation, the stimuli (A) which cause an act of speech and the reactions (C) which result from it, are very closely linked, because every person acts indifferently as speaker or hearer. We are free, therefore, without further discussion to speak of *vocal features* or sound (B) and of *stimulus-reaction features* (A-C) of speech.

From this point on, having once split the speech event, Bloomfield is confronted with *forms* (vocal features) on the one hand and *meanings*, (the correlated stimulus-reaction features of utterances) on the other hand. Now, since meanings (thus postulated) cannot be further dealt with, least of all analyzed into smaller units, then "This is no doubt why linguists, confronted with the parallelism of form and meaning, chose form as the basis of classification (**10**, p. 157)."

But leaving for the moment the problem of meaning, which bedevils Bloomfield continuously, we turn to his more general treatment (**11**) of language utterances, especially as it involves itself in psychological analysis. In principle, there is here nothing other than in preceding behavioristic systems. Note, for example, that "The language of any speech-community appears to an observer as a complicated signaling system . . . (p. 281)." "The phonemes . . . are the units of signaling (p. 136)." "A form is often said to *express* its meaning (p. 141).*

Markey. Since Markey's general theoretical standpoint stems from Mead (**123**), there is considerable concern with language as social process which gets its meaning when the hearer "takes the role of the other individual" (**121**, p. 401) but its substitute-stimulus value is indicated in the following definition: "A language habit is a word stimulus or action stimulus, which is a substitute for or symbolic of a bodily habit or process, and an absent stimulus. . . . The substitute character in regard to these two factors is an important consideration of the definition." Markey has elaborated his symbolic theory further within a somewhat freely-interpreted conditioned-response framework in *The Symbolic Process* (**122**).

---

* Note the familiar resemblance to the "form-content" problem in the mentalistic procedure and the faint traces of "expression." Here follows another nugget which serves to connect up Bloomfield's refurbished approach with the now forsaken expression theory: "The mental processes or internal bodily processes are known to each one of us only from speech-utterances and other observable actions (as if there were non-observable ones!). Since these are all we have to work with the mentalist in practice defines meanings exactly as does the mechanist (**11**, pp. 143–144).

Hunter. For Hunter (**83**) language is symbolic behavior which consists of responding to an absent object through a substitute stimulus of any sort. Its operation is similar to that of the kinesthetic cue for an animal in the delayed reaction experiment the function of which is that "of a *symbol*, because it enables the animal to react to an absent object in a selective manner, although it—the cue—has not been continuously present. This means that *we define a symbol as any process which is a substitute, which can be recalled if it ceases to be present* (p. 67)." The indirect nature of the response to the absent object is at once apparent in Hunter's formulation. One should also point out that included in language is *any* such indirect response, whether stimulated by an internal or external stimulus. Verbal behavior, one form of language, reaches its highest development in man and enables him "to behave with reference to absent objects in a way that makes science, religion, and the arts possible" (**83**, p. 330).

Dockeray. Dockeray (**40**, p. 142) follows Weiss in acknowledging language to be a form of stimulation of others as well as a response, and builds his machinery on a conditioned response base. A child, in the process of learning the language response, "milk," is described as follows: "The child sees milk; he hears the word, *milk*. He drinks milk and again hears the word, and so on, until *milk* has been definitely established as a symbol for a particular thing (p. 505)." There is here a first clear presentation of the symbolic relationship between word and absent object, which indicates the serial response of an individual to the word symbol, then to the object for which it stands—a relationship, it will be recalled, that Watson's formulation required but lacked.

Dashiell. According to Dashiell (**34**), language is a form of symbolic, social behavior in which words "stand for and represent" (p. 515) absent objects, thus extending man's behavioral space time. An example (Fig. 132, p. 512) shows Tom Sawyer signalling to Huck Finn, a symbolic stimulus which elicits a retreating response to the absent-but-fast-approaching-Aunt stimulus on the other side of the fence.

How the printed word "ORANGE" may be reacted to in a sign-and-thing signified relationship is described as follows:

> . . . reacting to the black-marks-on-white is not reacting to some larger whole of which they are a part, as would be the case in recognizing a car by its noise or a burning pudding by its odor or an orange by the feel of it in a Christmas stocking. The child is responding to something with which the pattern of black marks has been artificially associated and for which it then has come to function as a substitute. It is not a full and complete substitute, to be treated as the original object or situation would be treated; it is always handled as in-relation-to that other (p. 461).

Ruch. While no definition of language behavior is given, enough evidence of Ruch's (148) behavioristic leanings is given in a few scattered statements. Again, the child is said to learn the names of objects by a conditioning process.

> Eventually the child will be able to execute simple commands, such as "Bring me the doll from your basket." At this point we can definitely say that symbolic behavior is present. The child who hunts for the absent object is responding symbolically. The physically absent object is represented by some present symbol, and another symbol serves to direct the child toward the absent object (p. 618).

Carroll. In a recent issue of the *Psychological Review*, Carroll (24) has outlined a sign-theory of language with the intent of doing justice to the work already done in formal linguistics (*e.g.*, the speech unit, regularity, rate of occurrence, et cetera) and integrating these facts with systematic psychological theory. Toward this end, he adopts Morris' *semiosis* and extends his *pragmatics*. Peirce's *token* and *type* furnish models for *response* and *response type*, *response type* also being connected with Bloomfield's "speech community" (possessors of a cluster of "response-types").

Out of these components Carroll develops a symbolic theory which is best summarized in his own words.

> A linguistic response subsumed under a linguistic response-type must first of all play a role in a sign-situation such that "something takes account of something else mediately." Secondly, the sign-situation must be such that an organism deliberately intends another organism to respond to the sign *as* a sign. Consequently, the response must be one which can be produced either directly or indirectly by mechanisms.* Thirdly, the linguistic response must either itself be a sign, or it must form a segment of a sign. (This would include phonemes). Fourthly, there must be some social agreement among the members of a speech community with regard to the sign. Finally, the linguistic response-type must exhibit certain static and dynamic properties as a recurrent uniformity in the speech behavior of a number of individuals (p. 118).

Interbehavioral Language Theory

Rejecting all preceding Mentalistic (Body+Mind) and Behavioristic (Body−Mind) theories, Interbehavioral language construction builds within a general framework in which, except for degree of complexity, an organism's responses to stimulus objects are treated as *naturalistically* as gravitational and chemical interactions.

---

* Carroll would include within language such acts as an organism ringing a doorbell (p. 106), "the conventionalized choreographic movements of the modern dance" (p. 111) or "the deliberate wearing of one's hat at a rakish angle to suggest, say, the journalistic nonchalance of the city desk" (p. 112).

Language responses are segregated from other classes of behaviors when it is noted that the former are *bistimultational.* Viewed from the standpoint of the speaker, the organism concurrently *responds* to the hearer (the auxiliary stimulus) *and* the object talked about or referred to (the adjustment stimulus). On the side of the hearer, we see a similar bistimulational response involving the hearer's response to the *speaker's action,* which may be vocal, manual or other behavior (the auxiliary stimulus) and the thing talked about (the adjustment stimulus). There is no connection with theories that assume an "expression of ideas or thoughts" from one person to another; nor is the speaker's action *reified* into a series of symbols with which the hearer reacts *successively* and *which* refer *him* to objects (present or absent) as posited by both expression-symbol and behavioristic substitute-stimulus theories.

Furthermore, language is referential action in which the speaker *refers* to, and the hearer *is referred* to, some object, person or situation. Of course, an individual can talk to himself, or to put it another way, he can serve as his own auxiliary stimulus, and if he should talk to himself about himself, he also functions as the adjustment stimulus in such a situation. Typically, however, language is interpersonal and communicative.

> In studying communicative language we must differentiate between the transmissive and receptive forms. The former constitutes linguistic action which serves as a stimulus for the language action of another person. Receptive language, on the other hand, is a definite response to transmissive language, either of the vocal auditory type (language heard) or of the graphic visual form (language read) (**101**, pp. 301–302).

Since psychological language must necessarily constitute someone's behavior, this feature distinguishes it from non-psychological language. When a linguist deals with transcriptions of an individual's verbal patterning, he has a *symbolization* (alphabetic, phonetic, phonemic) of that individual's buccal action *at the time* he did perform a linguistic response; but, obviously, such a record is not *the response,* because when the individual performed that linguistic behavior, it was a bistimulational behavior segment that involved an auxiliary and an adjustment stimulus as was indicated above. This point is emphasized because of the common misconception that has lead to an identification of written records with linguistic behavior. Kantor (**100**, p. 305) has suggested the terms "morphological" and "functional" language to distinguish "thing" language from language behavior. For example, refined analysis (**102**, p. 451, **104**, p. 325) has lead to a distinction between symbolic behavior in which a verbal symbol (such as a calendar pad notation) serves as a substitute stimulus to elicit a response to an object present or absent. The individual's response in such an instance would be *serial* or *successive*—first to the substitute stimulus, then to its designatum or referent. The verbal nature of the note on the calendar pad would be incidental, too, since any kind of *object* often recalls someone or something. But the distinguishing feature between verbal symbolic behavior and linguistic responses is the *succession* of response of the former and the *concurrent* bistimulational factors in the latter.

How the Various Viewpoints Handle a Given Speech Occurrence

Space restrictions prevent an elaborate analysis of Interbehavioral language theory. It would, perhaps, be more to the point to conclude by taking a concrete, speech occurrence and to show how the three chief types of construction would describe such a behavior situation.

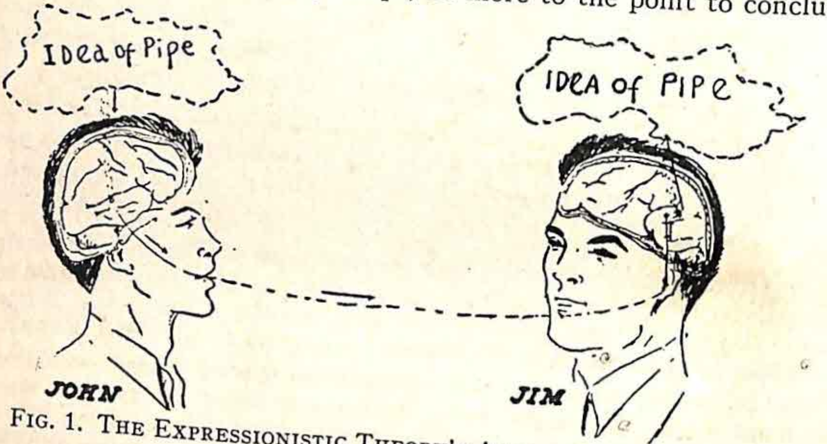


FIG. 1. THE EXPRESSIONISTIC THEORY'S ACCOUNT OF A LANGUAGE SITUATION

A language event in which John asserts, "There's your pipe." Diagrammed in accordance with expressionistic theory which maintains that the language event starts with certain ideas, desires, or feelings in John's "mind" which are transformed into neural impulses to the organs of speech. Then sound waves are propagated from John's mouth to Jim's ears by a "purely physical process." From this point on, the procedure is in inverse order to that occurring in John. There is first a physiological transmission of the acoustic image to the brain and finally a "psychical" association of this "image" with the corresponding "idea" of the pipe. Note that the series of events includes psychical-physiological-physical-physiological-psychical phases with gaps between each. (For other examples cf. Pillsbury E. Meader, Stinchfield, Gardiner, and Gray.)

Assume a situation in which John and Jim are customarily seated in the same room after dinner. Jim is looking for his pipe, John sees him searching for it and says, "There's your pipe" (indicating the pipe on a nearby coffee table).

A Mentalistic Version. According to the Expression theory, John sees Jim searching for his pipe (Fig. 1). Knowing where it is, he "wants to convey the idea" of its location to Jim. Immediately, certain associated word-images are aroused which start motor impulses that put mouth, tongue, laryngeal and other musculature into action that results in articulation. A series of sound waves passes from John's mouth to Jim's ears, where they are translated into neural impulses that go up to the brain, where further translation results in an arousal of an idea in Jim's mind that corresponds to the original one in John's mind. Fig. 1, is intended to bring out the never-to-be explained gaps in the chain of events which are alleged to be psychical→physiological→physical→physiological→psychical in turn.

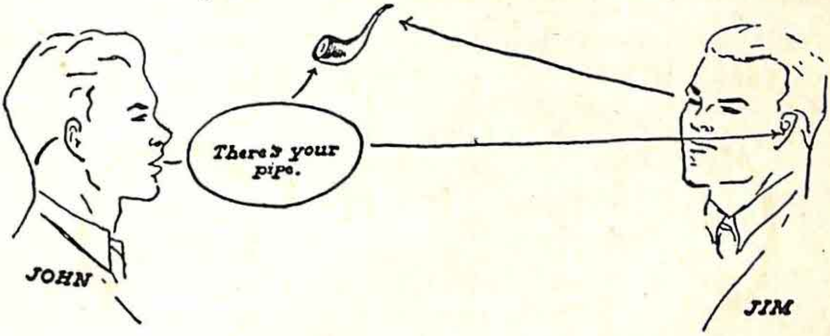
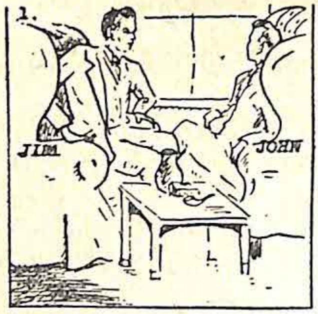


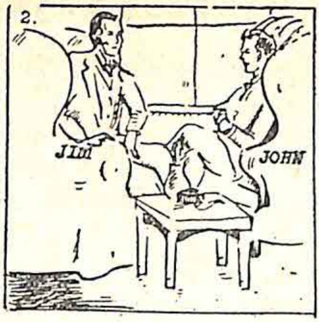
FIG. 2. HOW THE SYMBOLIC THEORY HANDLES A LINGUISTIC EVENT

The occurrence during which John addresses the remark, "There's your pipe." to Jim (same event as represented in Fig. 1) is here described as a symbolic situation. John's function is reduced to that of a *symbol-producer* and his speech *action* reified into a series of *things* that serve as stimuli for Jim. From this point on, the story may follow one of two courses: a. Behavioristic account—Jim reacts to the sounds then to the things for which they stand (i.e., the action involved is serial or successive). b. Mentalistic description—Jim reacts to the speech symbols, which arouse "thoughts" of the things to which they refer, which story would necessitate a diagram like that shown in Fig. 1. The symbols here are only *indirectly* connected with their referents via "the mind." (Exponents of a: Watson, Weiss, Exponents of b: Ogden & Richards, Sapir & others.)

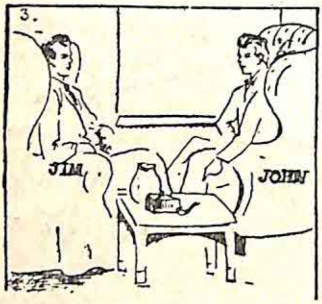
Another Mentalistic account (Fig. 2) would vary the account slightly from the preceding except as a matter of emphasis. With the expression phase assumed but not developed, it would *reify* John's action into a series of as-if-frozen symbols. In other words, John's role in the speech event is quite incidental, being that of a producer of symbols. Now, Jim reacts to these products or symbols (according to Sapir, even to such



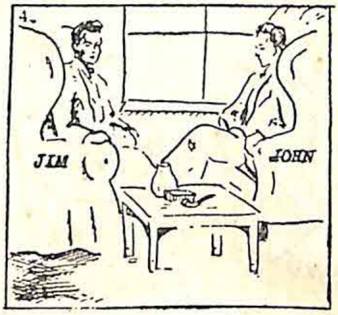
1. John looks out of window. Jim searches for his pipe. Unistimulational, non-linguistic behavior in both cases.



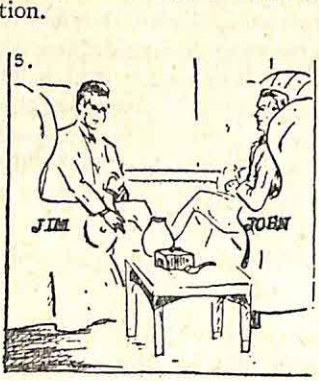
2. John sees Jim searching for his pipe. This illustrates interpersonal action on part of John but behavior of each is still unistimulational and non-linguistic.



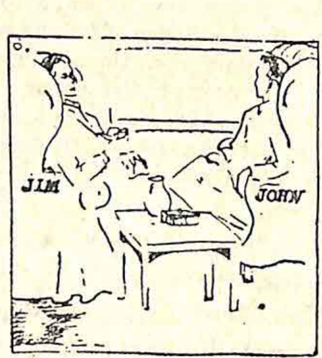
3. John says, "There's your pipe," (or he can merely point). In this case John performs a *bistimulational* (language) action because he is at the same time in contact with Jim and pipe. This is a *referor* action.



4. Illustrating Jim's *referee* language behavior. *Bistimulational* also because Jim interacts with John's speaking action and pipe in the same behavior segment.



5. Jim says, "Thanks," or he may grunt, "uh huh." The first is only morphological language; the second, verbal gesturing. Neither is bistimulation of referential.



6. John again looks out of window. Jim manipulates pipe (effective action). Behavior of each is unistimulational, non-linguistic.

FIG. 3. AN INTERBEHAVIORIAL DESCRIPTION OF A LANGUAGE ACT

The same even as that shown in Figs. 1 and 2 is here described but only the behavior segments shown in #3 and #4 constitute language behavior.

segments as prefixes and suffixes) which arouse "ideas" in his mind as above.

A Behavioristic Explanation. Fig. 2, may equally well illustrate the behavioristic account of the same speech happening since the psychic phases demanded by Mentalistic theories have been omitted in the diagram, but there are certain common features, namely, John's somewhat inferior role as symbol producer and Jim's serial action in his response to the symbols or signs (substitute stimuli) and then to the pipe for which they stand. Such accounts have been offered in an incipient form by Watson but have been more fully developed by Dashiell, Dockeray and Carroll among others.

An Interbehavioral Description. The same situation as suggested above permits an Interbehavioral analysis (Fig. 3) that rules out a number of behavior segments only related to the speech event but not themselves constituting language action.

1. If, for example, in Scene 1 John should be looking out of the window at a blooming rose bush, he would be performing unistimulational behavior (non-linguistic). The same would be true for Jim in his search of the pipe.

2. If, in turning from the window, (Scene 2) John sees Jim (still looking for his pipe), John is performing interpersonal behavior, but the action of both is still unistimulational and non-linguistic. John's action will be followed in the next behavior segment by a linguistic response but his action in the present behavior segment should not be confused or identified with it.

3. John says, "There's your pipe," or he points with his finger or nods in that direction with his head. In any case we have transmissive or referor action. This is the first language behavior segment to occur in the situation (Scene 3). It is a linguistic response on John's part because it is *bistimulational*—that is, John's behavior cannot be adequately handled without at the same time including Jim (auxiliary stimulus) and pipe (adjustment stimulus).

4. In Scene 4 Jim also performs a linguistic response because within the same behavior segment he has psychologically interbehaved with John's *action* (auxiliary stimulus) and to the pipe (adjustment stimulus). Jim's action is not reified into a symbol product, the inadequacy of which would be most apparent in the case of John's gesturing. It is pertinent to ask: for what could the gesture be a symbol?

5. In Scene 5 if Jim should say, "Thanks," this would be morphologically language but he would not, in this case, be referring to anything, nor would Jim from his side be referred to anything. Since it would not be bistimulational behavior, it would not be referential action but would more closely approximate a "verbal formula" or a conventionalized verbal gesture.

6. Assuming John to return to his gazing out the window and Jim to the manipulation of his pipe (an *effective* response) we would again in Scene 6 have unistimulational and therefore, non-linguistic behavior.

The tool which permits the differentiations of behavior indicated above is the "behavior segment" feature of Interbehavioral analysis.

As a result, behaviors showing a superficial resemblance at first blush show marked differences upon dissection. Furthermore, the inclusion of data covering the full range of linguistic behavior data as well as the coherence and precision of statement achieved within this framework contrast markedly with the sterility of previous theories. Finally, Interbehavioral language theory affords opportunity for the formulation of linguistic investigation that would go beyond the word count studies so endlessly inspired by the Expression and Symbol theories.

BIBLIOGRAPHY

1. Adams, S. Analysis of verb forms in the speech of young children and their relation to the language learning process. *J. exp. Educ.*, 1938, **7**, 141–144.
2. Adams, S. & Powers F. F. The psychology of language. *Psychol. Bull.*, 1929, **26**, 241–260.
3. Anderson, J. E. The development of spoken language. In **38***th Yearb. Nat. Soc. Stud. Educ. Part 1. Child development and the curriculum.* Bloomington, Ind.: Public School Publishing Co., 1939. Pp. 211–224.
4. Arsenian S. *Bilingualism and mental development.* Teach. Coll. Contr. Educ., No. 712. New York: Bur. Publ., Teachers College, Columbia Univ., 1937.
5. Bagley, W. C. The appreciation of the spoken sentence: a study in the psychology of language. *Amer. J. Psycho.*, 1900–01, **12**, 80–134.
6. Bartlett, F. C. & Smith, E. M. Is thinking merely the action of language mechanisms? (I) *Brit. J. Psychol.*, 1920, **11**, 55–66.
7. Bentley, M. & Varon, E. J. An accessory study of "phonetic symbolism." *Amer. J. Psychol.*, 1933, **45**, 78–86.
8. Bloomfield, L. *An introduction to the study of language.* New York: Henry Holt, 1914.
9. Bloomfield, L. Why a linguistic society. *Language*, 1925, **1**, 1–5.
10. Bloomfield, L. A set of postulates for the science of language. *Language*, 1925, **2**, 153–164.
11. Bloomfield, L. *Language.* New York: Henry Holt, 1938.
12. Boder, D. P. The adjective-verb-quotient: a contribution to the psychology of language. *Psychol. Rec.*, 1940, **3**, 310–343.
13. Brigance, W. N. The language learning of a child. *J. appl. Psychol.*, 1934, **18**, 143–154.
14. Briones, I. T. An experimental comparison of two forms of linguistic learning. *Psychol. Rec.*, 1937, **16**, 205–214.
15. Britton, K. *Communication, a philosophical study of language.* London: Kengan Paul, Trench, Trubner & Co., 1939.
16. Buchanan, M. A. & MacPhee, E. D. *An annotated bibliography of modern language methodology.* Toronto: Univ. Toronto Press, 1928.
17. Bucklew, J. An experimental set-up for the investigation of language problems. *J. exp. Psychol.*, 1914, **28**, 534–536.
18. Buhler, K. *Sprachtheorie: Die Darstellungsfunktion der Sprache.* Jena: Fischer, 1934.
19. Burger, H. & Kaiser, L. Speech without a larynx. *Acta. Oto-Laryngol.*, 1925, **8**, 90–116.
20. Carmichael, L., Hogan, H. P. & Walter, A. A. An experimental

study of the effect of language on the reproduction of visually perceived form. *J. exp. Psychol.*, 1932, **15**, 73–86.

21. Carroll, J. B. Diversity of vocabulary and the harmonic series law of word-frequency distribution. *Psychol. Rec.*, 1938, **2**, 379–386.
22. Carroll, J. B. Determining and numerating adjectives in children's speech. *Child Developm.*, 1939, **10**, 215–229.
23. Carroll, J. B. Knowledge of English roots and affixes as related to vocabulary and Latin study. *J. educ. Res.*, 1940, **51**, 102–119.
24. Carroll, J. B. The analysis of verbal behavior. *Psychol. Rev.*, 1944, **51**, 102–119.
25. Cassirer, E. Philosophie der symbolischen Formen. I. Teil, Die Sprache, 1923: II. Teil, Das mythische Denken. Berlin: Bruno Cassirer. 1925.
26. Cattell, J. M. The time it takes to see and name objects. *Mind*, 1886, **11**, 64.
27. Chapin, Mary V. & Washburn, M. F. A study of the images representing the concept "meaning." *Amer. J. Psychol.*, 1912, **23**, 109–114.
28. Chase, S. *The tyranny of words.* New York: Harcourt, Brace, 1928.
29. Cofer, C. N. & Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
30. Cofer, C. N., Janis, M. G. & Rowell, M. M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.
31. Cotton, J. C. Normal "visual hearing." *Science*, 1935, **82**, 592–593.
32. Critchley, M. *The language of gesture.* London, New York: Edward Arnold, Longmans Green, 1939.
33. Croce, B. *Aesthetic.* (Trans. by D. Ainslie). New York: Macmillan, 1909.
34. Dashiell, J. F. *Fundamentals of general psychology.* New York: Houghton Mifflin, 1937.
35. Davis, E. A. Development in the use of proper names. *Child Developm.*, 1937, **8**, 270–272.
36. Davis, E. A. Mean sentence length compared with long and short sentences as a reliable measure of language development. *Child Developm.*, 1937, **8**, 69–79.
37. Davis, E. A. The mental and linguistic superiority of only girls. *Child Developm.*, 1937, **8**, 139–143.
38. Davis, E. A. Developmental changes in the distribution of parts of speech. *Child Developm.*, 1938, **9**, 309–317.
39. Dewey, J. The psychology of infant language. *Psychol. Rev.*, 1894, **1**, 63–66.
40. Dockeray, F. C. *General psychology.* New York: Prentice-Hall, 1936.
41. Dockeray, F. C. *Psychology.* New York: Prentice-Hall, 1942.
42. Efron, D. *Gesture and environment.* New York: King's Crown Press, 1941.
43. Eisenson, J. *The psychology of speech.* New York: Crofts, 1940.
44. Eliason, N. E. & Davis, R. C. The effect of stress upon quantity in dissyllables. Indiana Univ. Publications, Science Series No. 8, 1939, 1–56.
45. Esper, E. A. A contribution to the experimental study of analogy. *Psychol. Rev.*, 1918, **25**, 468–487.
46. Esper, E. A. The psychology of language. *Psychol. Bull.*, 1921, **18**, 490–496.
47. Esper, E. A. A technique for the experimental investigation of associative interference in artificial lin-

guistic material. *Linguistic Soc. Amer., Lang. Monog.*, 1925, **1**, 1–47.

48. Esper, E. A. Studies in linguistic behavior organization: I. Characteristics of unstable verbal reactions. *J. gen. Psychol.*, 1933, **8**, 346–381.

49. Esper, E. A. Language. In *A handbook of social psychology*. Worcester, Mass.: Clark Univ. Press, 1935. Pp. 417–460.

50. Estes, W. K. A visual form of the verbal summator. *Psychol. Rec.*, 1940, **4**, 174–180.

51. Fenton, J. C. *A practical psychology of babyhood.* New York: Houghton Mifflin, 1925.

52. Fletcher, H. *Speech and hearing.* New York: Van Nostrand, 1929.

53. Foley, J. P., Jr. & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.

54. Foley, J. P., Jr. & Macmillan, Z. L. Mediated generalization and the interpretation of verbal behavior: V. "Free association" as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.

55. Foley, J. P., Jr. & Mathews, M. Mediated generalization and the interpretation of verbal behavior: IV. Experimental study of the development of interlinguistic synonym gradients. *J. exp. Psychol.*, 1943, **33**, 188–200.

56. Fox, C. W. An experimental study of naming. *Amer. J. Psychol.*, 1935, **47**, 545–579.

57. Fries, C. C. & Traver, A. A. *English word lists: a study of their adaptability for instruction.* Wash. D. C.: American Council on Education, 1940.

58. Gansl, I. Vocabulary: its measurement and growth. *Arch. Psychol., N.Y*, 1939, No. 236. Pp. 52.

59. Gardiner, A. H. *The theory of speech and language.* Oxford: Clarendon Press, 1932.

60. Garth, T. R. & Smith, O. D. The performance of full-blooded Indians on language and non-language intelligence tests. *J. abnorm. soc. Psychol.*, 1936, **32**, 376–381.

61. Goldstein, M. A. Speech without a tongue. *J. Speech Disorders*, 1940, **5**, 65–69.

62. Goodenough, F. L. The use of pronouns by young children: A note on the development of self-awareness. *J. genet. Psychol.*, 1938, **52**, 333–346.

63. Gray, L. H. *Foundations of language.* New York: Macmillan, 1939.

64. Gray, W. S. Summary of reading investigations. *J. Educ. Res.*, 1933, **26**, 401–424.

65. Grinstead, W. J. An experiment in the learning of foreign words. *J. educ. Psychol.*, 1915, **6**, 242–245.

66. Haas, Mary R. Types of reduplication in Thai. New Haven: *Yale Grad. School, Studies in Linguistics*, 1942, **4**, 1–6.

67. Hackman, R. B. & Duel, H. W. Do high school students who study a foreign language acquire larger vocabularies, spell their words more correctly and use better English than high school students who study no foreign language? *J. Amer. Ass. college Registr.*, 1941, **16**, 155–162.

68. Hagboldt, P. Physiological and psychological aspects of reading. In Hatfield, J. R., Leopold, W. & Zieglschmid, A. J. F., *Curme volume of linguistic studies.* Philadelphia: Linguistic Society of America, Language Monogr. No. 7, 1938. Pp. 84–91.

69. Hahn, E. F. *Stuttering: significant theories and therapies.* Stanford Univ., Calif.: Stanford Univ. Press, 1943.

70. Hall, Jr., R. A. The Romance words for flute. New Haven: Yale Grad. School., *Studies in Linguistics* 1942, **9**, 1–5.
71. Hargan, J. The psychology of prison language. *J. abnorm. soc. Psychol.*, 1935, **30**, 359–365.
72. Harrington, J. P. The American Indian sign language. *Indians at Work*, 1938, **5**, 8–13.
73. Harris, Z. & Voegelin, C. F. Hidatsa texts collected by Robert H. Lowie with grammatical notes and phonograph transcriptions. Indiana Historical Society. *Prehistory Research Series*, 1939, **1**, 1–227.
74. Hayakawa, S. I. *Language in action.* New York: Harcourt, Brace, 1941.
75. Henmon, V. A. C. The relation between mode of presentation and retention. *Psychol. Rev.*, 1912, **19**, 79–96.
76. Henry, J. The linguistic expression of emotion. *Amer. Anthrop.*, 1936, **38**, 250–256.
77. Herman, D. T. An harmonic analysis study of hoarse and non-hoarse voice quality. Bloomington; Indiana Univ., 1940.
78. Hill, H. E. An interbehavioral analysis of several aspects of stuttering. *J. gen. Psychol.* (In press).
79. Hockett, C. F. English verb inflection. New Haven: *Yale Grad. School, Studies in Linguistics*, 1942, **3**, 1–8.
80. Howells, T. H. An experimental study of speech perception. *Psychol. Bull.*, 1933, **30**, 690.
81. Howells, T. H. & Schoolland, J. B. An experimental study of speech perception. *J. gen. Psychol.*, 1934, **11**, 337–347.
82. Hudgins, C. V. & Di Carlo, L. M. An experimental study of assimilation between abutting consonants. *J. gen. Psychol.*, 1939, **20**, 449–469.
83. Hunter, W. S. *Human behavior.* Chicago: Univ. Chicago Press, 1928.
84. Hurlock, E. B. *Child development.* New York: McGraw-Hill, 1942.
85. Huse, H. R. *The psychology of foreign language study.* Chapel Hill: Univ. N. Carolina Press, 1931.
86. Huxley, A. L. *Words and their meanings.* Los Angeles: Ward Ritchie, 1940.
87. Irwin, O. C. & Chen, H. P. Speech sound elements during the first year of life: A review of the literature. *J. Speech Disorders*, 1943, **8**, 109–121.
88. Irwin, O. C. & Curry, T. Vowel elements in the crying vocalization of infants under ten days of age. *Child Developm.*, 1941, **12**, 99–109.
89. Jacobson, E. On meaning and understanding. *Amer. J. Psychol.*, 1911, **22**, 553.
90. Janus, S. Q. An investigation of the relationship between children's language and their play. *J. genet. Psychol.*, 1943, **62**, 3–62.
91. Jersild, A. T. *Child psychology.* New York: Prentice-Hall, 1941.
92. Jersild, A. T. & Ritzman, R. Aspects of language development: the growth of loquacity and vocabulary. *Child Developm.*, 1938, **9**, 243–259.
93. Jespersen, O. *Language, its nature, development and origin.* New York: Henry Holt, 1922.
94. Jespersen, O. *The philosophy of grammar.* London: Allen & Unwin, 1924.
95. Johnson, D. M. & Reynolds, F. A. A factor analysis of verbal ability. *Psychol. Rec.*, 1941, **4**, 183–195.
96. Johnson, W., et al. A study of the onset and development of stuttering. *J. Speech Disorders*, 1942, **7**, 251–257.
97. Judson, L. W. & Weaver, A. T. *Voice science.* New York: Crofts, 1942.
98. Kaiser, L. Examen phonétique expérimental d'un sujet privé de larynx. *Arch. neerl. de physiol.*, 1926, **10**, 468–480.

99. KANTOR, J. R. *Principles of psychology.* Vol. II. New York: Knopf, 1926.
100. KANTOR, J. R. *A survey of the science of psychology.* Bloomington, Ind.: Principia Press, 1933.
101. KANTOR, J. R. An objective psychology of grammar. Indiana Univ. Science Service, 1936, No. 1. Pp. 344.
102. KANTOR, J. R. The role of language in logic and science. *J. Philos.*, 1938, **35**, 449–463.
103. KANTOR, J. R. Postulates for a logic of specificity. *J. Philos.*, 1940, **37**, 29–42.
104. KANTOR, J. R. An interbehavioral analysis of propositions. *Psychol. Rec.*, 1943, **5**, 309–339.
105. KORZYBSKI, A. *Science and sanity: An introduction to non-Aristotelian systems and general semantics.* Lancaster, Pa.: Science Press Printing Co., 1941.
106. KROEBER, A. L. *Anthropology.* New York: Harcourt Brace, 1923.
107. KROUT, M. H. Autistic gestures. An experimental study in symbolic movement. *Psychol. Monogr.*, 1935, **46**, No. 208. Pp. 126.
108. KROUT, M. H. The social and psychological significance of gestures (a differential analysis). *J. genet. Psychol.*, 1935, **47**, 385–412.
109. KROUT, M. H. Understanding human gestures. *Sci. Mo.*, 1939, **49**, 167–172.
110. KUO, H. H. A study of the language development of Chinese Children. *Chinese J. Psychol.*, 1937, 334–364. (Abstr.).
111. LABRANT, L. L. A study of certain language developments of children in grades four to twelve, inclusive. *Genet. Psychol. Monogr.*, 1933, **14**, 387–491.
112. LEE, I. J. *Language habits in human affairs.* New York: Harper, 1941.
113. LUCKIESH, M. & MOSS, F. K. *Reading as a visual task.* New York: Van Nostrand, 1942.
114. LUKENS, H. Preliminary report on the learning of language. *Ped. Sem.*, 1894, **3**, 324–460.
115. MCCARTHY, D. The vocalization of infants. *Psychol. Bull.*, 1929, **26**, 625–651.
116. MCCARTHY, D. Language development. In Murchison, C., (Ed.) *A handbook of child psychology.* Worcester, Mass.: Clark Univ. Press, 1931. Pp. 278–315.
117. MCDAVID, R. I. Some principles for American dialect study. New Haven: *Yale Grad. School, Studies in Linguistics*, 1942, **12**, 1–11.
118. MCDONOUGH, A. R. The development of meaning. *Psychol. Monogr.*, 1919, **27**, Pp. 113.
119. MCGRANAHAN, D. V. The psychology of language. *Psychol. Bull.*, 1936, **33**, 178–216.
120. MATINOWSKI, B. The problem of learning in primitive languages. In OGDEN, C. K. & RICHARDS, I. A. *The meaning of meaning.* New York: Harcourt, Brace, 1938. Pp. 296–336.
121. MARKEY, J. F. The place of language habits in a behavioristic explanation of consciousness. *Psychol., Rev.*, 1935, **32**, 384–401.
122. MARKEY, J. F. *The symbolic process and its integration in children.* New York: Harcourt, Brace, 1928.
123. MEAD, G. H. A behavioristic account of the significant symbol. *J. Philos.*, 1922, **19**, 152–163.
124. MERRY, F. K. & MERRY, R. V. *From infancy to adolescence.* New York: Harper, 1940.
125. METCALF, J. T. Psychological studies of literary form. *Psychol. Bull.*, 1928, **35**, 337–357.
126. MITCHELL, A. J. The effect of bilingualism in the treatment of intelligence. *Elem. Sch. J.*, 1937, **38**, 29–37.

127. Montagu, H. Lip reading—a continuing necessity. *J. Speech Disorders*, 1943, **8**, 257–268.

128. Moore, T. V. The temporal relations of meaning and imagery. *Psychol. Rev.*, 1915, **22**, 177–215.

129. Morgan, J. J. B. *Psychology.* New York: Farrar & Rinehart, 1943.

130. Mosier, C. L. A psychometric study of meaning. *Psychol. Bull.*, 1939, **36**, 607–608.

131. Muenzinger, K. F. *Psychology: the science of behavior.* New York: Harper, 1942.

132. Nagge, J. W. *Psychology of the child.* New York: Ronald Press, 1942.

133. Newman, S. S. Further experiments in phonetic symbolism. *Amer. J. Psychol.*, 1933, **45**, 53–75.

134. Ogden, C. K. & Richards, I. A. *The meaning of meaning* (5th ed.) New York: Harcourt, Brace, 1938.

135. Olson, W. C. & Koetzle, V. S. Amount and rate of talking of young children. *J. exp. Educ.*, 1936, **5**, 175–179.

136. Paget, R. *Human speech.* New York: Harcourt, Brace, 1930.

137. Palmer, L. R. *An introduction to modern linguistics.* London: Macmillan, 1936.

138. Panconcelli, Calzia G. Boden und Sprache. *Volk. u. Rasse*, 1937, **12**, 385.

139. Piaget, J. *The language and thought of the child.* New York: Harcourt, Brace, 1926.

140. Pieron, H. *Thought and the brain.* New York: Harcourt, Brace, 1927.

141. Pillsbury, W. B. & Meader, C. L. *The psychology of language.* New York: Appleton, 1928.

142. Powers, F. F. Psychology of language learning. *Psychol. Bull.*, 1929, **26**, 261–274.

143. Preston, K. A. The speed of word perception and its relation to reading ability. *J. Gen. Psychol.*, 1935, **13**, 199–203.

144. Ragsdale, C. E. Language development in childhood. In Skinner, C. E. & Harriman, P. L. (Ed.) *Child Psychology.* New York: Macmillan, 1941. Pp. 129–133.

145. Reinecke, J. E. "Pidgin English" in Hawaii: a local study in the sociology of language. *Amer. J. Sociol.*, 1938, **5**, 778–789.

146. Richardine, Sister & Wilson, F. T. A reading activity in grade one. *Elem. Eng. Rev.*, 1938. **15**, 170–178.

147. Robinson, A. Is thinking merely the action of language mechanisms? (V) *Brit. J. Psychol.*, 1920, **11**, 81–86.

148. Ruch, F. L. *Psychology and life.* New York: Scott, Foresman, 1937.

149. Russell, B. *Philosophy.* New York: Norton, 1927.

150. Russell, G. O. *The vowel.* Columbus: Ohio State Univ. Press, 1928.

151. Sapir, E. A study in phonetic symbolism. *J. exp. Psychol.*, 1929, **12**, 225–239.

152. Sapir, E. *Language, an introduction to the study of speech.* New York; Harcourt, Brace, 1939.

153. Schlauch, M. Recent Soviet studies in linguistics. *Sci. and Soc.*, 1936, **1**, 152–167.

154. Scripture, E. W. *The elements of experimental phonetics.* New York: Scribners, 1904.

155. Scripture, E. W. Speech without using the larynx. *J. Physiol.*, 1915–16, **50**, 397–403.

156. Scripture, E. W. *Anwendung der graphischen methode auf Sprache und Gesang.* Leipzig: Barth, 1927.

157. Scripture, E. W. Macrophonic speech. *J. exp. Psychol.*, 1935, **18**, 748–791.

158. Sebeok, T. A. Phonemic system of Santali. *J. Amer. Oriental Soc.*, 1943, **63**, 66–68.

159. Seth, G. & Guthrie, D. *Speech in childhood.* London: Oxford Univ. Press, 1935.

160. Shirley, M. Common content in the speech of preschool children. *Child Developm.*, 1938, **9**, 333–346.

161. Skinner, B. F. The verbal summator and a method for the study of latent speech. *J. Psychol.*, 1936, **2**, 71–107.

162. Skinner, B. F. The distribution of associated words. *Psychol. Rec.*, 1937, **1**, 71–76.

163. Skinner, B. F. The alliteration in Shakespeare's sonnets: a study in literary behavior. *Psychol. Rec.*, 1939, **3**, 186–192.

164. Solomon, M. Stuttering as an emotional and personality disorder. *J. Speech Disorders*, 1939, **4**, 347–358.

165. Stark, W. A. The effect of bilingualism on general intelligence: an investigation carried out in certain Dublin primary schools. *Brit. J. Educ. Psychol.*, 1940, **10**, 78–79 (Abstr.).

166. Stetson, R. H. Speech movements in action. *Transactions of the American Laryngological Ass'n.*, 1933, **55**, 29–41.

167. Stetson, R. H. & Fuller, F. L. Diphthong formation, a preliminary study. *Arch. Neerl. de phonet. exper.*, 1930, **5**, 1–6.

168. Stinchfield, S. M. *The psychology of speech.* Boston: Expression Co., 1928.

169. Strayer, L. C. Language and growth: the relative efficacy of early and deferred vocabulary training, studied by the method of co-twin control. *Genet. Psychol. Monogr.*, 1930, **8**, 209–319.

170. Tharp, J. B. & McDonald, K. Psychology and methods in the high school and college: foreign languages. *Rev. Educ. Res.*, 1938, **8**, 35–38.

171. Thomson, G. H. Is thinking merely the action of laryngeal mechanisms. *But. J. Psychol.*, 1920, **11**, 67–70.

172. Thorndike, E. L. Experiments in euphony. *Psychol. Bull.*, 1934, **31**, 679–680.

173. Thorndike, E. L. On the number of words of any given frequency of use. *Psychol. Rec.*, 1937, **1**, 399–406.

174. Thorndike, E. L. Studies in the psychology of language. *Arch. Psychol., N. Y.*, 1938, No. 231. Pp. 67.

175. Thorndike, E. L. Some complications of associative processes. *J. exp. Psychol.*, 1943, **32**, 501–506.

176. Trager, G. L. The historical phonology of the Tiwa languages. New Haven: *Yale Grad. School. Studies in Linguistics*, 1942, **5**.

177. Travis, L. E. *Speech pathology.* New York: Appleton-Century, 1931.

178. Trussell, M. A. The diagnostic value of the verbal summator. *J. abnorm. soc. Psychol.*, 1939, **34**, 533–538.

179. Urban, W. M. *Language and reality: the philosophy of language and the principles of symbolism.* New York: Macmillan, 1939.

180. Voegelin, C. F. *Shawnee stems and the Jacob P. Dunn Miami Dictionary.* Indianapolis, Ind.: Historical Soc., 1938–40. Vol. I.

181. Vygotsky, L. S. & Luria, A. R. The function and fate of egocentric speech. In *Ninth Int. Cong. Psychol. Proc. and Papers.* Princeton: Psychol. Rev. Co., 1930. Pp. 464–465.

182. Walpole, H. R. *Semantics, the nature of words and their meanings.* New York: Norton, 1941.

183. Watson, J. B. *Behavior, and introduction to comparative psychology.* New York: Henry Holt, 1914.

184. Watson, J. B. *Psychology from the standpoint of a behaviorist.* Philadelphia: Lippincott, 1919.

185. Watson, J. B. *Behaviorism.* New York: Norton, 1930.

186. Watson, J. B. Behaviorism. Pages

327–329 in *Encyclopedia Britannica*, 1943, Vol. 3.

187. Wisenburg, T. & McBride, K. E. *Aphasia, a clinical and psychological study*. New York: The Commonwealth Fund, 1935.
188. Weiss, A. P. Linguistics and psychology. *Language*, 1925, **1**, 52–57.
189. Weiss, A. P. A theoretical basis of human behavior. Columbus, Ohio: Adams, 1929.
190. Weld, H. P. Meaning and process as distinguished by the reaction method. In *studies in psychology contributed by colleagues and former students of Edward Bradford Titchener*. Worcester, Mass.: Wilson, 1917. Pp. 181–208.
191. Whitaker, B. L. Speech-reading: a selected bibliography. *J. Speech Disorders*, 1943, **8**, 269–270.
192. Williams, C. B. A note on the statistical analysis of sentence length as a criterion of literary style. *Biometrika*, 1940, **31**, 356–361.
193. Wolfle, D. L. The relation between linguistic structure and associative interference in artificial linguistic material. *Linguistic Soc. Amer., Lang. Monogr.*, 1932, **11**, 1–55.
194. Woodworth, R. S. *Experimental psychology*. New York: Henry Holt, 1938.
195. Woolbert, C. H. The audience. *Psychol. Monogr.*, 1916, **21**, 36–54.
196. Worbois, G. M. Language development of children in two types of rural environments. *J. exp. Educ.*, 1942, **10**, 150–152.
197. Wundt, W. *Die Sprache*. 2 vols. Stuttgart: Kroner, (4). 1921.
198. Young, Florence M. An analysis of certain variables in a developmental study of language. *Genet. Psychol. Monogr.*, 1941, **23**, 31–143.
199. Young, Florence M. Development as indicated by a study of pronouns. *J. genet. Psychol.*, 1942, **61**, 125–134.
200. Yule, G. U. On sentence length as a statistical characteristic of style in prose; with application to two cases of disputed authorship. *Biometrika*, 1938, **30**, 363–390.
201. Zipf, G. K. *The psychobiology of language*. New York: Houghton Mifflin, 1935.

OBJECTIVE MEASUREMENT OF CLINICAL STATUS IN PSYCHOPATHOLOGICAL RESEARCH

DANIEL I. MALAMUD
Norwich State Hospital

What therapies are most effective for the various functional disorders? This problem has long been a pressing one in applied psychopathology, and will become even more important as new treatments are developed. Its solution requires careful experimental planning and the use of control groups and appropriate statistical techniques. Perhaps the most important requirement, if the relative values of different treatments are to be validly determined, is that treatments be investigated repeatedly with comparable groups and results be reported in comparable terms. Unfortunately, standard means for describing research populations and their degree of clinical improvement have not yet been established. As a consequence, experimental investigations of treatments have been handicapped. It is difficult for the investigator to build up equated research groups either for the purpose of determining the degree of superiority of a treatment over routine hospital care or for determining the relative values of two different treatments. The estimation of differences in clinical improvement between research groups is also beset with difficulties. Even the evaluation of a single treatment administered to a single diagnostic group by different investigators has become an uncertain task. Jessner and Ryan (36), for example, note that the statistics of insulin treatment as reported in the literature for schizophrenia vary from 70 per cent recovery to 20 per cent improvement, while some studies find even poorer results than in untreated cases. Such widely divergent results are difficult to reconcile so long as research groups, bearing the same descriptive labels, vary to an unknown extent in initial clinical status or subsequent degree of improvement.*

It seems that sound experimental studies of treatment in the functional disorders must first rest upon firm descriptive foundations. The present paper, therfore, will be devoted to

1. an analysis of present day description of clinical status in psychopathological research,
2. a tentative formulation of the salient features of a desirable descriptive system, and

---

* Other possible variables influencing these treatment figures and requiring control are (1) the mode of application of the therapy, (2) the influences of adjuvant therapy, (3) the period of follow-up study, and (4) the investigator's bias.

3. an evaluation of the role clinical psychology can play in contributing to the development of such a system.

PRESENT DAY DESCRIPTION OF CLINICAL STATUS

Perhaps the most basic type of study in treatment research is one designed to determine the degree of superiority of a therapy over routine hospital care. Such a study must meet at least two major requirements; first, its experimental and control groups must be comparable and second, its standards for assessing clinical improvement must coincide. Obviously, similar requirements would hold for studies designed to determine the relative values of two different treatments or when different investigators wish to compare validly results obtained with the same treatment. Let us examine these requirements more closely and the extent to which present day methods of describing clinical status help fulfill each of them.

Describing Research Populations

Experimental and control groups, to be comparable, must be matched as closely as possible for (1) capacity for spontaneous improvement (or prognostic outlook) and (2) qualitative status. The efficacy of a treatment is measured in terms of the degree to which improvement under this treatment exceeds the degree of spontaneous improvement to be expected with only routine care. Consequently, if experimental and control populations are to be compared with respect to recovery rates, they ought to be initially comparable in their capacity for spontaneous improvement. Also, if the value of a given treatment for a given type of patient is to be determined validly, both control and experimental groups must equally represent the qualitative status in question.

Capacity for spontaneous improvement. Various symptomatological features (process symptoms, atypicality, elements of confusion, etc.) and background factors (duration of illness, type of onset, pre-psychotic personality, etc.) are related to capacity for spontaneous improvement (**14, 54**). In most studies, however, the prognostic outlook of treated groups is defined solely in terms of diagnostic category, duration of illness, or both. Although diagnoses bear some relation to prognosis, their limited reliability and validity (to be discussed in the following section) make them unsuitable indices of capacity for spontaneous improvement. Duration of illness, the single background factor most commonly used for designating prognostic outlook, can be considered no more satisfactory. Although there is considerable agreement concerning the relation between duration and outcome of illness, this relation is insufficiently reliable in the individual case. For example, Hunt, Feldman and Fiero (32) note that no improvement took place in 45 per cent of 166 schizophrenics ill for less than six months, while improvement occurred in 25 per cent of 330 cases with more than 18 months duration. It should also be noted that the estimation of duration of illness, based as it is on information obtained from relatives and friends, is frequently undependable and at best only approximate.

Capacity for spontaneous improvement, then, cannot be satisfactorily assessed in terms of any one or two gross factors. Rather a numerically graded evaluation of the total complex of factors bearing significant relation to prognostic outlook seems required. Until these prognostic aspects of research populations are rigorously and uniformly defined, comparison of improvement rates must remain a dubious procedure.

Qualitative status. Research workers describe the qualitative status of patients under treatment chiefly in Kraepelinian diagnostic terms. Kraepelin's classification, although it originally tended to clarify the definitions of the functional syndromes, at present lacks sufficient *reliability* and *validity* for satisfactory coordination of research. Diagnostic criteria vary markedly from hospital to hospital and from one psychiatrist to another. For example, Boisen (5) comparing the relative number of schizophrenic subtypes in Illinois with hospitals in other sections of the country found that whereas in Massachusetts 30% of the newly admitted schizophrenics are classed as catatonics, Illinois has only 2.7%. On the other hand, whereas Illinois places 45% in the hebephrenic group, Massachusetts finds only 15.5%. Myerson (47) points out that even when psychiatrists agree as to what constitutes a specific mental disorder they may still make inconsistent diagnoses because of the difficulties in correctly recognizing criteria symptoms (e.g., negativism, resistiveness, apathy) when they occur in individual cases. The consequent unreliability of diagnoses limits to an unknown extent the comparability of patient groups, particularly those treated by different investigators.

A description of the patient's qualitative status, to be most valid, must include all the patient's clinical trends and their relative intensities. Kraepelinian diagnoses, however, are not sufficiently refined for such comprehensive description. Functional psychoses frequently represent a mixture of schizophrenic and affective elements. Yet many research workers, treating the major functional psychoses as if they were demarcated entities each with its own pathology, ignore secondary clinical trends and force the case being considered into one or the other major classification. They then consider, at least for research purposes, all cases under the same label as constituting a homogeneous disease entity.* Referring to this procedure, Cameron (**10**) states:

> Most of the confusion and the contradictions in the literature on functional psychoses can be traced to the wholly unjustifiable practice of treating the heterogeneous mass of major depressive and manic disturbances as if they really constituted a single illness, Kraepelin's "manic-depressive insanity," and of lumping together the schizophrenic disorders likewise as a homogeneous disease entity as though it were as uniform as chicken-pox.

The tendency to classify patients into one discrete category or another obscures not only the qualitative mixtures in clinical trends but also their relative degrees of manifestation. For example, one patient diagnosed "schizophrenia, catatonic type" may exhibit marked catatonic features, moderate paranoid trends and traces of hebephrenic symptoms, while another similarly classified patient, although equally catatonic, may be less paranoid and more hebephrenic. Gross, ungraded diagnostic terms conceal these sometimes significant differences in symptom intensity. Yet investigations have clearly indicated that individual differences in most psychological traits are distributed on a graded and continuous scale. Undoubtedly, most psychopathological trends are similarly distributed. Terms already exist for describing gross degrees of mania and depression, but these are rarely used to describe research groups. Furthermore, no equivalent terms of gradation exist for schizophrenia. Obviously, lumping patients together under the same diagnostic label who vary in the relative strengths of their clinical trends tends to obscure further their qualitative status.

Describing Clinical Improvement

If recovery rates are to be validly compared, the same standards of clinical improvement must be applied to the different research groups under consideration. Unfortunately, the improvement categories customarily employed vary in use and meaning from one investigator to another. The term "improved," for example, has been used in one study (**60**) to mean symptom amelioration plus ability to adjust in the outside community, while in another investigation (**48**) it was applied to patients showing symptom amelioration *without* the ability to adjust on the outside. In some studies (**48**) the patient is classified as "recovered" if he returns to his prepsychotic status, while in others it is also required that he develop insight regarding his illness (16, 58). Some psychiatrists do not consider any of their patients capable of recovering from schizophrenia (60).

---

* It is recognized that many research workers are well aware of the shortcomings of this procedure and consider it a necessary convenience. Although the author sympathizes with the need for this expediency, it must be realized that such a procedure represents a barrier to future progress.

Judgments of clinical improvement lack not only a standard terminology but also objective methods of derivation. Their reliability and validity are therefore doubtful. Cohen, Malmo and Thale (17) appear to recognize this when they state:

> The lack of acceptable objective methods of evaluating the nature and changes of activity in mentally ill patients is probably one of the greatest obstacles to the valid interpretation of psychiatric data. The usual clinical judgments of "improved," "unchanged," or "worse" are too frequently derived from anecdotal data, offered by ward nurses and attendants. It is true that much ward personnel has the inestimable advantage of observing given patients for prolonged periods. But it must be admitted that, failing checked objective methods of observation, their opinion (necessarily a composite one) may be formed in a quite capricious fashion.

A valid statement of changes in intensity of disorder would obviously depend upon adequate descriptions of qualitative status both before and after treatment. Unfortunately, the difficulties that apply to present day description of initial qualitative status obtain equally in describing changes in status.

OUTLINE OF A DESIRABLE DESCRIPTIVE SYSTEM

From the foregoing analysis it seems clear that many of the difficulties involved in the evaluation of a treatment's efficacy can be traced largely to the lack of objective, standard descriptions of clinical status. This lack stems from the emphasis of present day descriptive methods upon crude, discontinuous categories instead of well-defined, measurable variables. True, the rough qualitative mapping out of the functional syndromes was a necessary stage in psychopathology, but the time now seems ripe, and the need urgent, for attempting a transition to more precise, objective and possibly more fundamental modes of description. Such a desirable descriptive system would involve the following procedures:

1. The functional psychoses would be described by a minimal number of clearly defined, qualitative components (or dimensions) of abnormality.*

---

* It is assumed that such basic variables can be isolated. Studies supporting this assumption will be presented in a later section. Future research must decide whether these components would be identical with present day syndromes or with more fundamental underlying variables.

2. These primary components would be elicited and measured by objective tests.
3. Analysis of clinical status would require measurement of each component.
4. Both capacity for spontaneous improvement and qualitative status would be described in terms of *cross-sectional* patternings of the same component strengths.*
5. Clinical improvement would be described in terms of *longitudinal* changes in these patterns.

In such a multidimensional system, the descriptive questions would not be merely: "Is the patient schizophrenic or manic-depressive, or is he in a partial or full remission?" Rather they would become: "*To what extent does the patient exhibit each of the basis components of functional disorder, and how do his present component strengths compare with what they were before treatment?*" This comprehensive mode of description would automatically indicate capacity for spontaneous improvement, better provide for "mixed" syndromes and relative intensity of clinical trends, and would lay a more accurate and reliable basis for defining clinical improvement.

Clinical psychology, long concerned with the problems of exact measurement, may be of considerable aid in identifying and measuring primary components of abnormality. Psychological testing and methods of "psychostatistical" analysis appear to be particularly promising and appropriate tools for attacking this problem. The essential features of psychological testing are

1. the devising of standard situations considered representative of broader areas of behavior,
2. the quantitative measurement of individual differences in reactions to these situations, and
3. the comparison of the individual's performance with previously established norms.

By *psychostatistical* analysis we refer to those factorial methods which have been developed for deriving from many measures the minimal number of variables necessary for description of behavior. In the following section we shall survey a number of psychological studies containing significant hints as to how these and other approaches may contribute to the development of a more useful and objective descriptive schema.

---

* From an ahistorical view, only contemporary qualitative status can determine capacity for spontaneous improvement. Background factors are of significance only insofar as they indicate the present condition of the patient. From this point of view, objective tests are legitimate and probably the most desirable means of assessing capacity for spontaneous improvement. In the course of developing the above system, however, background factors will for some time probably supplement objective tests.

PSYCHOLOGICAL MEASUREMENT OF CLINICAL STATUS

Psychiatrists have long recognized the need for exact methods of description. According to Hollingworth (28), Kraepelin himself was much concerned with replacing vague, subjective judgments by precise, objective records of performance elicited under controlled conditions. Although he recognized the significance of the more complex aspects of personality, he preferred to limit his studies to functions he could measure, such as work capacity, learning rate, retentivity, etc. Since the Wundtian psychology of Kraepelin's time, however, advances in testing theory and technique have opened up the prospects of extending Kraepelin's admirable experimental program to include the deeper, more complex aspects of abnormal behavior. A vast literature has accumulated dealing with the application of psychological techniques to the psychiatric disorders. Although differing widely in procedures, aims and theory, these studies have generally emphasized exactness and objectivity. A comprehensive summary of most of these investigations can be found in Hunt (30), Hunt and Cofer (31), and Brody (8). The bulk of these psychological studies can be classified as follows:

1. Studies concerned with predicting outcome of shock therapy and bearing on the problem of assessing capacity for spontaneous improvement.
2. Studies in which attempts are made to differentiate qualitative types of disorder.
3. Studies mainly concerned with distinguishing groups varying in intensity of disorder.
4. Studies bearing on the problem of describing clinical status economically by means of basic variables.

Below we shall consider the most promising studies of each type and the implications they carry for the problems of this paper.*

Determining Capacity for Spontaneous Improvement

Patients suffering from the functional psychoses differ in their capacity for spontaneous improvement. Can these differences in capacity be assessed by psychological tests? In the absence of studies bearing directly on this question, let us examine related investigations attempting to determine capacity for positive response to shock treatment:

---

* For many of the studies one might question whether the results would have been the same if they had been carried out by other investigators using different diagnostic criteria. Despite this qualification, it is felt that the total body of results strongly supports the contention that a more objective, exact system of clinical definition and measurement can be derived by means of psychological techniques.

1. A number of investigators have found the Rorschach test useful in predicting response to shock treatment. Halpern (23) administered the Rorschach to 17 schizophrenics before insulin therapy. According to the test results, the group contained two types of patients. One group presented a picture of rigid personality and general nonproductiveness, incapable of emotional response and showing little capacity for identification with others. These individuals responded poorly to treatment. The other group, representing patients who displayed receptivity to emotional stimuli and a capacity for empathy, attained satisfactory adjustment for a period of a year or more. The three Rorschach signs differentiating these two groups would have enabled correct prediction of treatment response in 14 out of 17 cases.

On the basis of "blind" analyses of Rorschach records, Piotrowski (51) was able to predict improvement resulting from insulin treatment correctly 88.3% of the time, i.e., in 53 out of 60 cases. With added experience he found that more precise definition of prognostic signs would have yielded correct predictions in 93.3% of the cases, i.e., in 97 of 104 patients. He concluded that the greater the degree of intellectual regression the less the chances of improvement, while the greater the degree of emotional regression the better the chances of improvement. He felt that the Rorschach distinguished between these two types of regression with greater precision than would have been possible without its use.

Eisner and Orbison (18) administered the Rorschach to catatonics and manics before metrazol treatment, once under normal conditions and then again in the sodium amytal twilight state. They found that certain effects of sodium amytal on Rorschach performance were of prognostic significance. For example, the drug's affect-releasing action, as reflected in increased color scores, was greater in the more improved than in the less improved groups.

Morris (46) found six Rorschach signs which, if applied to selection of patients for metrazol treatment, would have enabled him to yield an improvement rate of 78%, a gain of 34% over clinical selection alone.

2. Concept formation tests have also been found to be of prognostic value. Bolles, Rosen and Landis (6) report that performance of 19 patients on a modified Vigotsky test of concept formation, the Weigl test and the BRL sorting test showed closer relation to outcome of insulin treatment than did the duration of illness.

Zubin and Thompson (76) found a correlation of .68 between pre-treatment test performance on the Weigl, BRL and Vigotsky and immediate outcome, and a correlation of .50 between combined background factors and immediate outcome in a group of 49 patients treated with insulin shock. If both criteria had been applied in the case of the insulin patients and only those patients who obtained favorable ratings on both had been selected for therapy, the probable improvement rate for immediate outcome would have been 95.4% and for outcome on follow-up 76.3%.

3. Combinations of objective tests have been found useful. Wechsler, Halpern and Jaros (71) found that a test battery consisting of a vocational interest blank, tests of counting by 3's and naming words in three minutes, a similarities and a directions test not only corresponded well with psychiatrists' estimates of effects of treatment, but also gave a correlation of .73 and a percentage of correspondence of 87 with a clinical appraisal of the patient's condition 6 to 18 months after treatment.

4. *Attempts at predicting treatment response by means other than objective tests have also been successful.* With various symptoms and background factors as prognostic criteria, Chase and Silverman (**15**) assessed the probable outcome for each of 150 patients treated with metrazol or insulin shock. These assessments were compared with later results of treatment. Of 43 cases in which the prognosis was considered good, 33 (77%) showed remissions. On the other hand, of 74 cases in which a poor outcome was predicted, 63 (85%) remained unimproved.

The foregoing type of statistical prediction may be considerably facilitated by use of a rating scale such as that devised by Wittman (**75**) which weights the various symptoms and background items considered prognostic. According to Wittman, this scale gave a definitely more accurate prediction of outcome following shock therapy than the average medical staff member's judgment.

The above studies carry the following implications for our problem:

1. Capacity for spontaneous improvement can probably be assessed with a relatively high degree of objectivity and quantification.

2. Development of an objective index of capacity for spontaneous improvement would require the establishment of a test battery correlating with spontaneous changes in clinical status rather than with responses to specific treatments. The mode of analysis necessary for establishing such a battery would be identical with that employed by the above studies, i.e., by finding the test items in the improved group which originally differentiated them from the unimproved group.

3. Only a negligible number of psychometric techniques have been studied from a predictive viewpoint. It appears likely that many of the tests to be suggested in the following sections for eliciting and measuring qualitative deviation would also have prognostic value.

4. It seems likely that certain drugs, such as sodium amytal, modify test performance in such a manner as to disclose existing trends more sharply or reveal hitherto latent capacities. Administering psychological tests to patients under the influence of such drugs, following the method of Eisner and Orbison (**18**), may increase the prognostic value of these tests considerably.

Determining Qualitative Types of Disorder

There are different qualitative syndromes among the functional psychoses. Do these various types of psychotics behave distinguishably on psychological tests? Do they show qualitatively different performance patterns? Let us examine the available evidence:

1. *Projective techniques have been useful in revealing the qualitative character and organization of various functional syndromes.* Klopfer and Kelley (**40**) sum-

marize the characteristic patterning of Rorschach signs which have been found for schizophrenic, depressed and hypomanic states. The fine differentiation of schizophrenic subgroups is still in the beginning stage " . . . although these types may be recognized in the individual case." The Thematic Apperception Test enables identification of psychiatric diagnoses with fair accuracy (24). The Szondi technique, in which the subject picks those he most likes and dislikes from six series of photographs of abnormal and criminal types, seems to have good potentialities (53). Angyal (3) finds that schizophrenics, reproducing tachistoscopically exposed letter combinations, reveal differences among themselves which are related to hebephrenic and paranoid characteristics.

2. *From Lewinian experiments comes objective evidence of the schizophrenic's apathy.* According to Hunt and Cofer (31), Stumbur found that unlike Zeigarnik's normals, schizophrenics failed to recall more "uncompleted" than "completed" tasks. Rickers-Ovsiankina (55) failed to find schizophrenics showing the normal tendency to resume "uncompleted" tasks. Escalona (19) found that normals, manics and depressed patients shifted characteristically in their level of aspiration under the effects of success and failure.

3. *Studies of prompted drawings of psychotics in terms of such characteristics as gestalt properties, color, movement and restraint have been found indicative of various functional syndromes.* These investigations have been well summarized by Anastasi and Foley (1, 2). Mira (44) claims he can deduce various diagnostic characteristics when changes of direction and magnitude are introduced into a subject's drawings as he tries to copy from a model.

4. *Temperamental factors isolated by factor analysis of groups of simple tests have been found to be related to functional syndromes.* Possibly the most systematic attempt at objective testing of abnormality has been undertaken by a group of English experimentalists under the leadership of Spearman (69). By means of factor analysis of groups of simple tests requiring a minimum of intelligence, they have isolated a number of temperamental factors ("p" or perseveration, "f" or fluency, etc.) which in combination with "g" have been found to differentiate between normals and various diagnostic groups. Characteristic combinations of "g," "p" and "f" test scores have been found for manics, depressives, hysterics, and schizophrenics.

5. *Inventories and questionnaires have been adapted to the discrimination of clinical syndromes.* The Humm-Wadsworth Inventory (29) and the Minnesota Multiphasic Scales (26) are outstanding examples of the questionnaire type of test which has been adapted to the discrimination of clinical syndromes and carefully validated by means of control groups.

6. *Analysis of abnormal patterns and scatter on standard intelligence tests appears promising.* Brody (8) and Hunt and Cofer (31) summarize a variety of psychometric studies designed to discover characteristic abnormal patterns of performance on standard intelligence tests. Recently Schafer and Rapaport (61) have outlined several new conceptions of scatter which appear to give clear expression to significant variations in the Wechsler-Bellevue score patterns of different clinical groups. Although no reliable generalizations can yet be made, many of these studies appear very promising.

7. *A variety of other measures of behavioral functioning or adaptation have interesting possibilities.* Reversal of the "fatigue effect" tendency with continuous work, slowness of perceived fluctuations in ambiguous figures and possibly

certain alterations of the patellar reflex have been pointed out by Hunt and Cofer (31) as specifically characteristic of manic-depressives. Studies of behavioral correlates of the moods of elation and despondency (e.g., speed and quality of associations, number of spontaneous remarks, time to make decisions) offer interesting possibilities for objective measures of mania and depression (20, 37). Hausmann (27) reports a method for objectively measuring "thinking difficulties." He believes this method can eventually be used for differentiating depressive and schizophrenic reactions at an early stage.

8. *Wertham* (73) *summarizes many ingenious tests, yet to be validated, which have been invented for discriminating cycloids from schizoids.* By means of these tests it has been found, for example, that cyclothymics are more sensitive to color than to form, and are more distractible than schizothymics.

The above studies carry the following implications for our problem:

1. The development of tests for eliciting and measuring qualitative syndromes of functional abnormality is still in its infancy and appears to be a difficult but not impossible task.

2. A wealth of experimental techniques have shown varying degrees of promise. The crying needs of most of these techniques are for validation, fuller standardization and more objective scoring schemes. Many need to be repeatedly applied to groups of patients covering the whole range of functional disorders and correlated with ratings of clearly defined clinical characteristics rather than with psychiatric diagnoses.

3. The most promising tests are those which disguise their true purpose, tap central regions of personality, permit multi-scoring, and require the subject to perform a specified task rather than report about himself.

4. A battery of tests, rather than any single measure, is most likely to be effective in qualitative diagnosis. Such a test battery would probably require (a) the establishment of critical scores for each subtest, i.e., points at which qualitatively different groups were maximally discriminated from each other, and (b) the development of diagnostic score-patterns.

Determining Quantitative Degree of Disorder

We have seen that the functional syndromes can probably be differentiated qualitatively by means of objective tests. Now we turn to the problem of intensity of disorder. If we conceive of normals, neurotics and psychotics as falling at different points on a normality-abnormality continuum, can test scores correctly place these different groups on this continuum? This is perhaps the simplest demand that can be made of psychological tests, i.e., that they grossly differentiate groups of normals from groups of patients varying widely in degree of mental illness, regardless of type. It is obvious that most of the tests listed in the foregoing section can probably meet this requirement, since they elicit not only different qualitative aspects of behavior but also measure their manifested intensity by means of graded scores. Below are listed additional studies which have centered specifically on this problem:

1. Normals, neurotics and psychotics, when ranked according to their weighted scores on a test battery which included the Rorschach, Spearman factor tests, Bernreuter, and Wells word association, were found to be in approximately the same relative position as that determined by clinical appraisal (41).

2. Cattell (11, 13) claims that his Cursive Miniature Situations test "... probably distinguishes between normals and psychotics more completely than any other test, motor or verbal, yet tried for that purpose." This test requires subjects to circle and cross out specified marks on a paper strip moving beneath a small aperture.

3. Penrose and Myers (49) have found that abnormal score patterns on a group intelligence test ("General Examination-M") distinguished between psychotic and non-psychotic subjects in about 75 per cent of the cases. They suggest this test pattern method as a screening device for identifying psychiatric suspects.

4. The Multiple Choice Rorschach (25), Rorschach "instability" scores (59), word association tests (38, 43, 72), Bernreuter test profiles (62), and a food check list (70) have also been found to discriminate to some degree between normals and various deviating groups.

5. Two *normal* groups representing extremes in camp adjustment were satisfactorily differentiated by a composite profile derived from three "stress" tests (57). Obviously, tests which distinguish normal groups varying in adjustment are even more likely to distinguish between more extreme deviant groups.

6. Tests of deterioration (4, 65, 66) and the more peripheral functions such as reaction time (34), tapping rate (63), and tachistoscopic perception (3) have also been found to be associated with judged severity of disorder. Rodnick and Shakow (56) studied reaction time to regular and irregular preparatory intervals of various lengths "... as a quantitative measure of the ability of the schizophrenic to reach and maintain a high level of preparation in meeting recurrent environmental stimulation." A composite index based on differences between patients and normals separated the two groups with practically no overlap.

7. Psychological tests have also discriminated between degrees of intensity of illness in the *same group tested at different times*. Differences between Rorschach responses (33, 39, 50), perseveration (67), concept formation (76), and intelligence test patterns (22, 35) when compared before and after treatment have shown positive correspondence with clinical improvement. A battery of tests including the Rorschach and tests of intelligence, perseveration, orientation, concept formation, word association, etc. showed a "fair correlation" between the clinical psychiatric rating and psychometric rating, before, during and after insulin shock therapy (42).

The above studies carry the following implications for our problem:

1. Reliable psychological tests for measuring intensity of functional

disorder can probably be developed within the near future. Such quantitative measures, even if gross, would represent a considerable advance over present day methods of evaluating clinical improvement.

2. Tests confronting subjects with an *unpredictable sequence of stimuli and continual necessity for maintaining set and attention* appear to be particularly effective means for measuring gross intensity of illness. Examples of such tests are Rodnick and Shakow's (56) reaction-time set-up with irregular warning procedures and Cattell's Cursive Miniature Situation (**11**, **13**), involving a continuously developing series of events in unpredictable succession.

3. Since no single test completely differentiates normal and disordered groups, a *battery* of tests tapping several levels of personality organization would probably be most successful in measuring intensity of illness.

4. Degrees of abnormality can probably best be discriminated by means of empirically derived *composite indices*, combining several weighted variables and analagous to those devised by Rodnick and Shakow (56) and Rodnick, Rubin and Freeman (57).

5. If tests are to be reliably applied to individuals, future studies must pay particular attention to the practical question of *overlap* between groups as well as to the significance of differences between group means.

Determining Basic Variables of Abnormality

The investigator who hopes to derive an exact and systematic description schema for assessing clinical status is confronted at the outset with a bewildering variety of clinical symptoms. Developing objective tests for each of these symptoms would be an almost insuperable task. Therefore, he may reasonably inquire: How can abnormality in the functional psychoses be described most *economically*? Are some symptoms more basic than others? Or do certain fundamental variables exist which underly all these diverse symptoms?

The various methods of factor analysis may provide a fruitful approach to these problems. These technical methods of correlational analysis aim to derive that minimal number of variables most adequate for the systematic and orderly description of personality. It is not within the province of this paper to discuss the pros and cons of the various systems of factor analysis, nor of the basic psychological and mathematical assumptions of the approach itself. Some of the more promising studies, however, which have adopted this general mode of analysis will be briefly described below:

1. *Orthodox factor analysis.* In a pioneering study, Moore (**45**) applied factor analysis to the intercorrelations of 40 ratings of psychiatric symptoms in 402 psychotics. He was able to extract 5 factors, which he called "catatonic,"

"deluded," "manic," "cognitive defect," and "constitutional hereditary depression." The possibility of determining basic syndromes by means of mathematical techniques was thus established.

Line and Griffin (**41**), in a search for basic measurable variables underlying mental stability, administered a large battery of tests to 43 subjects ranging in "stability" from mental hospital patients to graduate students. Analysis of the results indicated two factors in the scores on these tests, "objectivity" and "fluency." The investigators concluded: "If other factors are detected, and if they prove to be pertinent to the field in which we are interested, patterns of these factors as represented in the functionings of individuals may then be compared. Should these in any way correspond to clinical types, we may thus arrive at the point where factor-patterns give direction to the search for what constitutes the difference between types."

Cattell (**12**) has performed the largest scale factor analysis of personality measures to date. He has derived twelve primary factors which he has interpreted with reference to psychiatric syndromes. Although too early to judge, this monumental work may mark the first step towards the establishment of the type of descriptive system deemed desirable in this paper.

2. *Cluster or syndrome analysis* represents a more empirically obvious method of correlational analysis insofar as it merely aims to discover what correlation coefficients cluster together, i.e., intercorrelate positively among themselves and show negative correlations with all or nearly all other traits. It is assumed that these clusters represent true trait unities. By means of this mode of analysis Sheldon (**64**) has isolated three basic components of temperament (each represented by clusters of 20 traits), and has demonstrated their close correlation to the individual's basic morphological pattern. The relations between these components and the major functional disorders are being investigated.

3. Stephenson (**68**) has suggested an "*inverted factor technique*" for deriving personality types. This approach consists of obtaining the correlations between the scores of persons instead of tests. The factors obtained are not traits of personality but types of individuals, i.e., typical orderings of traits to which all persons of a type approximate. It is possible to give to each individual a score which indicates his degree of correlation with this "type-factor." Bordin (**7**) points out that this type of analysis permits more objective identification of factors than orthodox factor analysis: "Since one is now dealing with persons who have high and low factor loadings, the facts can be obtained in a public manner. Persons have many characteristics which are observable by all and on which all can agree. For example, they are male or female, they have been good or poor achievers, have lived in high level or low level economic environments, etc. These characteristics may then be related to the factor loadings by appropriate statistical procedures for indicating relationships and thus result in an identification or non-identification of factors. In essence, it represents a means whereby the best parts of the clinical and statistical methods may be combined."

The above studies carry the following implications for our problem:

1. With the aid of "psychostatistical" methods the derivation of basic components of abnormality appears to be possible.

2. This goal may be fulfilled by means of a systematic research

program which, in its simplest outlines, would involve the following steps: (a) Testing a large representative group of functional psychotics, upon admission and also after six months of routine hospital care, with a test battery sampling a wide range of behavior variables. (b) Rating this patient group before and after the six-month period on a large number of symptoms, ward behavior items and background factors.* (c) Analyzing "psychostatistically" the network of functional relations existing among tests and ratings.

3. The coordinated design of the above research plan would permit not only investigation of basic variables underlying the intercorrelations between tests and ratings, but also supplementary attacks upon the problems of objective measurement of prognostic outlook and of clinical improvement.†

By employing psychological tests and "psychostatistical" methods in organized research programs a frame of reference may be eventually derived within which functional disorders could be classified and scaled. Once fundamental dimensions of deviation are discovered (be they those of Moore (**45**), Line and Griffin (**41**), Cattell (**12**), or others), we may then standardize objective tests for eliciting and measuring each of these dimensions in any given patient. Each patient would thus be described by a set of scores assigning him his quantitative status with respect to each variable. Such a descriptive system would conceive of functional abnormalities not as discrete psychiatric categories, but as patterns of relatively independent, numerically graded components.

The program outlined above will obviously require many years of concerted effort on the part of both psychiatrists and clinical psychologists for fulfillment. Yet the problems attacked are of such a fundamental nature that the time and effort would undoubtedly be well spent. As sound descriptive foundations are established, the frequencies of successes and failures for different treatments will be stated and compared with ever greater confidence. The values of various therapies will be more adequately determined. The way will be open to the establishment of statistical prediction tables listing in ascending order the efficacy of treatments for different types of patients—an achievement which will raise the practice of applied psychopathology to a new level of effectiveness.

---

* A number of suggestive rating scales of clinical behavior have appeared (**9, 52, 74**) but their use in psychiatric research has been all too infrequent. The author agrees with Sheldon (**64**), particularly for research in psychopathology, " . . . that in our struggles with the complex data of psychology we have neglected unduly the technique of ordering, classifying and scaling our materials through ratings of trained judges."

† The experimental program outlined above is only a tentative proposal. New developments in the field or individual circumstances may require appropriate modifications in this design of research.

SUMMARY

1. Present day methods of describing clinical status lack satisfactory reliability and validity. The comparison and evaluation of treatment results will remain extremely difficult as long as research groups, bearing the same descriptive labels, vary to an unknown extent in their capacity for spontaneous improvement, qualitative status and degree of clinical improvement.

2. A desirable descriptive system would be characterized by a minimal number of abnormality components which could be elicited and measured by means of objective tests. Patients would be described in terms of their patternings of component strengths and shifts in these patterns.

3. Relevant studies involving psychological tests and "psycho-statistical" methods have been surveyed with a view to evaluating their potential contribution to the development of such a system.

BIBLIOGRAPHY

1. Anastasi, Anne & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal: II. Approaches and interrelationships. *Ann. N. Y. Acad. Sci.*, 1941, **42**, 1–112.
2. Anastasi, Anne & Foley, J. P., Jr. A survey of the literature on artistic behavior in the abnormal: IV. Experimental investigations. *J. gen. Psychol.*, 1941, **25**, 187–237.
3. Angyal, Alice F. Speed and pattern of perception in schizophrenic and normal persons. *Character & Pers.*, 1942, **11**, 108–127.
4. Babcock, Harriet. *Time and the mind.* Cambridge, Mass.: Sci-Art, 1941.
5. Boisen, A. T. Types of dementia praecox—a study in psychiatric classification. *Psychiatry*, 1938, **1**, 233–236.
6. Bolles, Marjorie M., Rosen, G. P. & Landis, C. Psychological performance tests as prognostic agents for the efficacy of insulin therapy in schizophrenia. *Psychiat. Quart.*, 1938, **12**, 733–737.
7. Bordin, E. S. Factor analysis in experimental designs in clinical and social psychology. *Psychol. Rev.*, 1943, **50**, 415–429.
8. Brody, M. B. Mental testing. *J. ment. Sci.*, 1944, **90**, 127–151.
9. Buck, J. N. The JNB Psychograph. *J. appl. Psychol.*, 1943, **27**, 65–74.
10. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Vol. II. Pp. 861–921.
11. Cattell, R. B. An objective test of character-temperament. *J. gen. Psychol.*, 1941, **25**, 59–73.
12. Cattell, R. B. Interpretation of the twelve primary personality factors. *Character & Pers.*, 1944, **13**, 55–91.
13. Cattell, R. B. An objective test of character-temperament. II. *J. soc. Psychol.*, 1944, **19**, 99–114.

14. Chase, L. S. & Silverman, S. Prognostic criteria in schizophrenia. *Amer. J. Psychiat.*, 1941, **98**, 360–368.
15. Chase, L. S. & Silverman, S. Prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1943, **98**, 464–473.
16. Cheney, C. O. & Clow, H. E. Prognostic factors in insulin shock therapy. *Amer. J. Psychiat.*, 1941, **97**, 1029–1039.
17. Cohen, L. H., Malmo, R. B. & Thale, T. Measurement of chronic psychotic over-activity by the Norwich Rating Scale. *J. gen. Psychol.*, 1944, **30**, 65–74.
18. Eisner, E. & Orbison, W. D. An objective evaluation of metrazol therapy: a Rorschach study. Paper presented before the annual meeting of the Amer. Psychiat. Assoc. Chicago, May, 1939.
19. Escalona, Sybille K. The effect of success and failure upon the level of aspiration and behavior in manic-depressive psychoses. *Univ. Ia. Stud. Child Welf.*, 1940, **16**, 199–302.
20. Fisher, V. E. & Marrow, A. J. Experimental study of moods. *Character & Pers.*, 1934, **2**, 201–208.
21. Franz, S. I. The time of some mental processes in the retardation and excitement of insanity. *Amer. J. Psychol.*, 1906, **17**, 38–68.
22. Graham, Virginia T. Psychological studies of hypoglycemia therapy. *J. Psychol.*, 1940, **10**, 327–358.
23. Halpern, Florence. Rorschach interpretation of the personality structure of schizophrenics who benefit from insulin therapy. *Psychiat. Quart.*, 1940, **14**, 826–833.
24. Harrison, R. Studies in the use and validity of the Thematic Apperception Test with mentally disordered patients: II. A quantitative validity study. III. Validation by the method of "blind analysis." *Character & Pers.*, 1940, **9**, 122–133; 134–138.
25. Harrower-Erickson, Mary R. & Steiner, M. E. *Large scale Rorschach techniques.* Springfield, Ill.: Charles C Thomas, 1945.
26. Hathaway, S. R. & McKinley, J. C. A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, **10**, 249–254.
27. Hausmann, M. F. A method to objectively demonstrate thinking difficulties. *Amer. J. Psychiat.*, 1933, **13**, 613–625.
28. Hollingworth, H. L. *Abnormal psychology.* New York: Ronald Press, 1930.
29. Humm, D. G. & Wadsworth, G. W., Jr. The Humm-Wadsworth Temperament Scale. *Amer. J. Psychiat.*, 1935, **92**, 163–200.
30. Hunt J. McV. Psychological experiments with disordered persons. *Psychol. Bull.*, 1936, **33**, 1–58.
31. Hunt, J. McV. & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944. Vol. II. Pp. 971–1032.
32. Hunt, R. C., Feldman, H. & Fiero, R. P. "Spontaneous" remissions in dementia praecox. *Psychiat. Quart.*, 1938, **12**, 414–425.
33. Hunt, Thelma. Personality profile studies. In W. Freeman. & J. W. Watts, *Psychosurgery.* Springfield, Ill.: Charles C Thomas, 1942. Pp. 164–181.
34. Huston, P. E., Shakow, D. & Riggs, L. A. Studies of motor function in schizophrenia: II. Reaction time. *J. gen. Psychol.*, 1937, **16**, 39–82.
35. Jastak, J. Psychometric changes following insulin therapy. *Delaware St. med. J.*, 1939, **11**, 114–119.
36. Jessner, L. & Ryan, V. G. *Shock treatment in psychiatry.* New York:

Grune & Stratton, 1941.

37. JOHNSON, WINIFRED B. Euphoric and depressed moods in normal subjects. *Character & Pers.*, 1937, **6**, 79–98.
38. KENT, GRACE H. & ROSANOFF, A. J. A study of association in insanity. *Amer. J. Insanity*, 1910, **67**, 37–126
39. KISKER, G. W. A projective approach to personality patterns during insulin-shock and metrazol-convulsive therapy. *J. abnorm. soc. Psychol.*, 1942, **37**, 102–124.
40. KLOPFER, B. & KELLEY, D. M. *The Rorschach technique.* Yonkers-on-Hudson, New York: World Book, 1942.
41. LINE, W. & GRIFFIN, J. D. M. The objective determination of factors underlying mental health. *Amer. J. Psychiat.*, 1935, **91**, 833–842.
42. MCNEEL, B. H., DEWAN, J. G., MYERS, C. R., PROCTOR, L. D. & GOODWIN, J. E. Parallel psychological, psychiatric and physiological findings in schizophrenic patients under insulin shock treatment. *Amer. J. Psychiat.*, 1941, **98**. 422–429.
43. MALLER, J. B. *Controlled association test.* New York: Bureau of Publications, Teachers College, Columbia Univ., 1936.
44. MIRA, E. Myokinetic psychodiagnosis: A new technique of exploring the conative trends of personality. *Proc. roy. Soc. Med.*, 1940, **33**, 173–194.
45. MOORE, T. V. The essential psychoses and their fundamental syndromes. *Stud. Psychol. & Psychiat.*, 1933, **3**, No. 3, Pp. 128+x.
46. MORRIS, W. W. Prognostic possibilities of the Rorschach method in metrazol therapy. *Amer. J. Psychiat.*, 1943, **100**, 222–230.
47. MYERSON, A. Medical psychiatry. In M. Bentley & E. V. Cowdry, *The problem of mental disorder* (1st Ed.). New York & London: McGraw-Hill, 1934. Pp. 28–38.
48. PACELLA, B. L. & BARRERA, S. E. Follow-up of a series of patients treated by electrically induced convulsions and by metrazol convulsions. *Amer. J. Psychiat.*, 1943, **99**, 513–518.
49. PENROSE, L. S. & MYERS, C. R. A method for the preliminary psychiatric "screening" of large groups. *Amer. J. Psychiat.*, 1941, **18**, 238–242.
50. PIOTROWSKI, Z. A. Rorschach manifestations of improvement in insulin treated schizophrenics. *Psychosom. med.*, 1939, **1**, 508–526.
51. PIOTROWSKI, Z. A. The Rorschach method as a prognostic aid in the insulin shock treatment of schizophrenics. *Psychiat. Quart.*, 1941, **15**, 807–822.
52. PLANT, J. S. Rating scheme for conduct. *Amer. J. Psychiat.*, 1922, **1**, 547–572.
53. RAPAPORT, D. The Szondi test. *Bull. Menninger Clin.*, 1941, **5**, 33–39.
54. RENNIE, T. Prognosis in manic-depressive psychoses. *Amer. J. Psychiat.*, 1942, **98**, 801–814.
55. RICKERS-OVSIANKINA, MARIA A. Studies on the personality structure of schizophrenic individuals: II. Reaction to interrupted tasks. *J. gen. Psychol.*, 1937, **16**, 179–196.
56. RODNICK, E. H. & SHAKOW, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, **97**, 214–225.
57. RODNICK, E. H., RUBIN, M. A. & FREEMAN, H. Related studies on adjustment. *Amer. J. Psychiat.*, 1943, **99**, 872–880.
58. ROSS, J. R. The pharmacological shock treatment of schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 769–779.
59. ROSS, W. D. A quantitative use of the

Rorschach method. *Amer. J. Psychiat.*, 1944, **101**, 100–104.

60. RUPP, C. & FLETCHER, ELIZABETH K. A five to ten year follow-up study of 641 schizophrenic cases. *Amer. J. Psychiat.*, 1940, **96**, 877–888.
61. SCHAFER, R. & RAPAPORT, D. The scatter in diagnostic intelligence testing. *Character & Pers.*, 1944, **12**, 275–284.
62. SCHMIDT, H. O. & BILLINGSLEA, F. Y. Test profiles as a diagnostic aid: the Bernreuter Inventory. *J. abnorm. soc. Psychol.*, 1945, **40**, 70–76.
63. SHAKOW, D. & HUSTON, P. E. Studies of motor function in schizophrenia: I. Speed of tapping. *J. gen. Psychol.*, 1936, **15**, 63–106.
64. SHELDON, W. H. & STEVENS, S. S. *The varieties of temperament.* New York: Harper, 1942.
65. SHIPLEY, W. C. & BURLINGAME, C. C. A convenient self-administering scale for measuring intellectual impairment in psychotics. *Amer. J. Psychiat.*, 1941, **97**, 1313–1324.
66. SIMMINS, C. Studies in experimental psychiatry: IV. Deterioration of "G" in psychotic patients. *J. ment. Sci.*, 1933, **79**, 704–734.
67. STEPHENSON, W. Studies in experimental psychiatry: III. p-score and inhibition for high p-praecox cases. *J. ment. Sci.*, 1932, **78**, 908–928.
68. STEPHENSON, W. Some recent contributions to the theory of psychometry, *Character and Pers.*, 1936, **4**, 294–304.
69. STEPHENSON, W., MACKENZIE, M., SIMMINS, C. A., KAPP, D. M., STUDMAN, G. L., & HUBERT, W. H. DE B. Spearman factors and psychiatry. *Brit. J. med. Psychol.*, 1934, **14**, 101–135.
70. WALLEN, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, **40**, 77–81.
71. WECHSLER, D., HALPERN, FLORENCE & JAROS, EUGENIA. Psychometric study of insulin-treated schrizophrenics. *Psychiat. Quart.*, 1940, **14**, 466–476.
72. WELLS, F. L. Autistic mechanisms in associated reactions. *Psychol. Rev.*, 1919, **26**, 376–381.
73. WERTHAM, F. Progress in psychiatry. IV. Experimental type psychology. *Arch. Neurol. Psychiat., Chicago*, 1930, **24**, 605–611.
74. WILCOX, P. H. The Gardner Behavior Chart. *Amer. J. Psychiat.*, 1942, **98**, 874–880.
75. WITTMAN, PHYLLIS. A scale for measuring prognosis in schizophrenic patients. *Elgin Pap.*, 1941, **4**, 20–33.
76. ZUBIN, J. & THOMPSON, J. *Sorting tests in relation to drug therapy in schizophrenia.* New York: New York State Psychiatric Institute, 1941, 1–23.

A HISTORY OF THE DEPARTMENT OF PSYCHOLOGY AT THE UNIVERSITY OF CHICAGO*

FORREST A. KINGSBURY
The University of Chicago

When the new University of Chicago opened its doors on October 1st, 1892, there was not, nor was there for ten years thereafter, a Department of Psychology. In the Department of Philosophy, later known as the Department of Philosophy and Education, several courses in psychology were offered; and of its first 14 Ph.D. dissertations, four dealt with psychological topics. In 1900 the first doctorate in Education was awarded, and three years later, in 1903, the newly independent Department of Psychology granted its first doctorate, to John Broadus Watson.

I. BEGINNINGS OF THE DEPARTMENT: HOUSING AND EQUIPMENT

Instruction in psychology, the first year of the University, was given in Cobb Hall, then the only lecture hall on the campus. After the erection of Ryerson Physical Laboratory in 1893–94 two rooms were provided for a psychological laboratory on the third floor, which were occupied for three years. The first description of its equipment mentions models of brain and sense-organs, and apparatus for vision, hearing, touch, and movement, and for time-measurement. With the erection in 1898 of the four Hull Biological Laboratories, the psychological laboratory was transferred to space on the first and fourth floors of the Anatomy Building, where it remained until 1901. In that year a two-story frame house at 5704 Ellis Avenue, located just north of where Ricketts North stands, was secured as the first exclusive home of what, in the following year, became the separate Department of Psychology. In this small edifice, supplemented by the Culver Hall basement room where Watson did his animal research work, all class, laboratory, research, and office work was carried on. Overcrowded from the beginning, this building was retained as the Department's animal laboratory for 21 years after 1908. In that year the other activities were moved to the more commodious quarters which the Department has occupied ever since, and Mr. Angell's description of "The New Psychology Building at the University of Chicago" in the November 1908 *Psychological Bulletin* makes entertaining reading to the present generation of students and visitors, whose admiration for these facilities seems more moderate than that formerly expressed. Nevertheless, later generations of students have agreed with a remark attributed to Mr. Angell, that "more psychology has been learned in the third-floor research rooms than in the first-floor lecture rooms."

---

* Paper read before the Psychological Journal Club, May 17, 1945. I am indebted to Professor Emeritus Harvey A. Carr for many helpful suggestions and criticisms in the course of its preparation.

One innovation was the provision of the Department's own machine shop in which much of the apparatus since used in Departmental research and instruction has been designed and constructed. In his 1902 report to the President (which, framed, hangs on the wall of the Departmental office), Mr. Angell stated a principle which has ever since been consistently adhered to: "In the securing of apparatus there has never been any purpose to create a museum in which specimens should be gathered for mere exhibit."

Another development was the Psychological Library. Psychological books and journals were originally housed in the general library in Cobb. In the 5704 Ellis building a few volumes were kept in the office, their number gradually increasing until a year or so after the Department had moved into its present quarters. Then the psychological journals were brought in and a more rapid growth began. The northeast second floor room became the Department library, one of a score or more on the quadrangles, with an overflow into the *Seminar Room* now occupied by the Department office. In the 1920's, this room, and later still another were added to the library, which by now was being increasingly used by readers from other Departments and schools, a reflection of psychology's widening affiliations and applications. Gradually the less used volumes were crowded off the shelves into unsuitable storerooms. Finally, in 1944, partly as a result of the trend toward centralization of libraries, partly due to the unsatisfactory, crowded condition of the collection and the constant fire-hazard, the staff reluctantly consented to the removal of the approximately 10,000 volumes to Harper, leaving a small reference collection in the laboratory for staff use, but freeing space for a more adequate Departmental office and seminar room.

A year after Mr. Lashley joined the Department in 1929 he was given for an animal laboratory the quarters in Culver Hall now occupied by Mr. Kluever. This continued to serve the Department until a year or two after Lashley's departure to Harvard in 1935. Later, the present quarters in Belfield were secured for Mr. Gulliksen's animal work. Meanwhile, Mr. Thurstone's psychometric laboratory had developed as a result of enlarged staff, improved equipment and widening areas of research. Increasing student enrollment in recent years has forced several Departmental classes into larger classrooms and laboratories than our building affords; so that the Department is now geographically quite dispersed.

Before the Department was separately organized, instruction in psychology was given by several members of the combined Department. The first year, psychological courses were offered by Charles A. Strong, Associate Professor of Psychology, who with several of his colleagues had come from Clark University to join the new University's staff, and by James Hayden Tufts, Assistant Professor of Philosophy, from the University of Michigan. That first year there was no full Professor or Head of the Department. In the first Register of the University we find Mr. Strong announced as offering an introductory course, using James' new *Briefer Course*, and a one-year advanced course providing "A critical study of James' Principles of Psychology with occasional lectures"; while Mr. Tufts offered an intermediate course using James' and Dewey's Psychologies, "with lectures and demonstrations." What was demonstrated is suggested by the course description in the following year, which reads: "Will be presented partly in the form of lectures, illustrated by experimental demonstrations; and each student will be given an opportunity to dissect an ox's eye and a sheep's brain." No one can doubt that psychology at the University of Chicago started off true to the Jamesian pattern.

II. ESTABLISHMENT OF THE LABORATORY: THE ANGELL PERIOD

By the second year, the idea of psychology as an experimental science was so thoroughly accepted that experimental courses, a laboratory, and a new laboratory director were announced. This was, therefore, one of the earliest psychological laboratories established in an American University. The director, Assistant Professor James Rowland Angell, had been a young instructor at the University of Minnesota, with Bachelor's and Master's degrees from Michigan, and a Master's at Harvard under James, followed by a year at Berlin and Halle. He returned to take the Minnesota position without waiting to receive a Ph.D. degree, except for an honorary one many years later. (He also received 18 LL.D's.) His studies under Dewey at Michigan and James at Harvard equipped him to be the chief psychological spokesman for that distinguished group of philosophers, Dewey, Tufts, Mead, Moore, and Ames, who came to be known as the *Chicago School* of philosophical pragmatists and psychological functionalists. Angell was Director of the Psychological Laboratory from 1894 to 1902, and thereafter Head of the Department until 1920. For twelve years he was Senior Dean or Dean of the Faculties of Arts, Literature, and Science, and in 1918–19 acting President of the University. In 1920 he resigned to become President of Carnegie Corporation and a year later President of Yale. Since retirement in 1937 he has continued to be active as Educational Counselor for the National Broadcasting Company.

An interesting and effective teacher, who stimulated his students to independent and critical thinking, Angell could, at the same time, appreciate diverse and even conflicting viewpoints and give them a fair showing. Consequently, his introductory *Psychology*, published in 1904, which supplanted James' *Briefer Course* as the most widely used text in the new psychology, presents alternative views so objectively that on certain issues it is difficult or impossible to state with assurance his own position. Nevertheless, the text, like his lectures, presents clearly a *functionalist* point of view, with its stress upon mental activities as operations, not as contents; and with emphasis on their biological mechanisms, their social relationships, and practical utilities. This catholicity of view characterized not only the wide variety of courses he offered, but the research that came from the laboratory during his term as director and chairman. The 39 doctoral dissertations of the Department's first 18 years, as well as the four earlier theses mentioned before, represent a wide range of interests—sensory, perceptual, attention, thought, and imagery functions, motor learning, and the newer fields of animal, social, and differential psychology. It was characteristic of this atmosphere that doctors from this Department are to be found in almost every field and division of psychological work, and represent every viewpoint —except, perhaps, some of the more extreme ideologies.

Because of his growing preoccupation with administrative duties in the later years of his service, the actual direction of the laboratory devolved increasingly upon Mr. Carr, so that even in 1915 when I had my first contact with Angell as a teacher, his occasional visits to the laboratory were events rare enough to be remarked.

III. TEACHING STAFF: THE DEWEY INFLUENCE

Not only from Angell, but from Dewey did the new psychological program get its direction. In 1894, two years after the opening, John Dewey came to Chicago from a professorship of Philosophy at Michigan, to become Head of the Department of Philosophy and Education, remaining for 10 years, when he went to Columbia University. The foremost figure in America, outside of James, in the new philosophical school of pragmatism (or, as he preferred to call it, *Instrumentalism*) he early became accustomed to being a storm-center of controversy, sometimes personal, oftener academic, but always a deliberate and dispassionate storm-center. While his influence on the psychology program operated principally through his associates and especially Angell, he did, throughout the ten years of his stay, offer a course in the Psychology of Ethics. This later found publication in the widely used Dewey and Tufts *Ethics*, and still later in his *Human Nature and Conduct.* His primary interest was in education and the development of a new-type laboratory school, described in his small but influential book *School and Society*. His views on the social goals and psychological methods of education have doubtless influenced American educational thought and practice more than those of any other educator.

During the first ten years, several other members of the combined Department, besides Dewey and Tufts, offered courses in psychology, among them George H. Mead, who taught comparative psychology, psychological methodology, and social psychology, Addison W. Moore, one of Angell's early assistants in the laboratory, Warner Fite who taught both laboratory and theoretical courses, and E. S. Ames, psychologist of religion. All these men except Fite remained for many years members of the Philosophy staff and their courses continued to attract some psychology students even after the Department's separation. Other names not so familiar to the present generation I shall not mention.

The second period, the 18 years of Mr. Angell's chairmanship, brought several staff members whose names are familiar to many of today's students. In 1903 John B. Watson received the first doctorate awarded in the Psychology Department. Following this he served on the staff for five years, going in 1908 to Hopkins as Professor of Experimental and Comparative Psychology. I have heard Mr. Carr say that Watson's introductory psychology class usually spent the first 9 or 10 weeks of the quarter on the nervous system. Even at that time he was enthusiastic over the idea of a psychology based on the reflex arc; although his first behavioristic pronouncement did not appear until 1912, after he had discovered Pavlov and the Conditioned Reflex.

The third doctorate in the Department was earned in 1905 by Harvey Carr who three years later returned as Assistant Professor and was identified with the Department until his retirement 30 years later. He told me that his first job was to supervise the remodeling of the new laboratory. Of his later work we shall speak presently. During this period, the Department's first teaching staff consisted of Angell and Watson, who were joined in 1907 for two years by Dr. Karl T. Waugh, and the following year by Carr. During this period, Professors Mead, Moore, and Ames of the Philosophy Department continued to contribute to Psychology programs, as did also Thomas of Sociology, Gore and later on Judd, Freeman, and Stevens of Education, and perhaps others who might be mentioned. Joseph W. Hayes, a 1911 Chicago Ph.D., now deceased, was instructor for several years in the second decade of the century, as were also Ellsworth Faris, later Professor of Sociology, and Harry D. Kitson, now of Teachers' College; while during the first World War years and immediately thereafter, the staff included Curt Rosenow, J. R. Kantor, now of Indiana, and R. T. Wiltbank, later at Knox College.

IV. THE CARR PERIOD

If we call the ten years in the Department of Philosophy the first period, and the years of Angell's chairmanship the second, the third period, that from 1920 to 1939, may be characterized as the period of Carr's leadership; even though for the first six years he was merely Director of the Laboratory, the chairmanship being held for four years by Charles H. Judd, Director of the School of Education, and for two years by James H. Tufts, Chairman of the Philosophy Department. Their influence, however, was almost altogether administrative, for only the former actually taught a course in the Department. While Carr's well known interests, in space perception, memory, and animal and human learning, are apparent in many of the doctoral theses of that period, several were written under the supervision of other men in the Department, as well as a few in other Departments, notably Education and Physiology. Later, after Professors Thurstone and Lashley joined the staff, psychometric and psychobiological studies became numerous and important. Their own contributions are too well known to call for description or evaluation here. In 1926 Carr was appointed Chairman of the Department, a position he retained until his retirement in 1938. In 1925 appeared his *Psychology*, perhaps the clearest presentation we have of the later, more biological formulation of a functionalist point of view. Carr himself, however, has always protested against being tagged with any particular label, even *functionalist*, holding that the views he stands for have too much in common with those held by others to justify thus setting them into contrast. This distaste for stereotyped classifications and rules characterized also his attitude of encouraging independence and self-reliance in his students. His scrutiny and criticism of research proposals and reports was always searching and detailed, but he was not dictatorial or dogmatic in his criticisms. He would ask, "What do you propose to do about this difficulty?" or "Have you thought of that possibility?", while appeals for advice were characteristically met by the reply, "It's up to you; do what *you* think best." It is not surprising that more than a hundred doctors and perhaps as many masters, whose dissertations he supervised or criticized, recall him with admiration and affection as a teacher who did his best to make them independent and resourceful scholars, not disciples.

This period saw the coming and going of a good many young teachers of introductory sections, most of them with fellow's or assistant's rank, and most of them later taking their doctorates. Up to a few years ago it was believed that introductory psychology could not properly be taught except in small sections; and the post-war years had brought a large influx of undergraduates. Some of these names would be unfamiliar to most readers, although some will be recognized as having later achieved some eminence in their fields, such as Blatz, Heron, Darrow, Sheldon, and others perhaps as well known as these. Other staff members of higher rank handled advanced courses and supervised research, among them the late Edward S. Robinson, later of Yale, Florence Richardson (later Mrs. Robinson) who carried on the psychological work in the School of Business, Arthur Kornhauser also of the School of Business, Arthur G. Bills, now of Cincinnati, Capt. H. A. Swenson, for long advisor in the College and instructor of introductory courses, K. S. Lashley, L. L. Thurstone, M. W. Richardson, H. Kluever, Harold Gulliksen, and the writer. All of these I have named (except Lashley and Kluever) were doctors of this Department, which fact occasionally used to lead to comments within the Departmental staff about the "dangers of academic inbreeding"; although, knowing the diversity of views and interests among those trained in the Department, I have never worried much about this hazard.

The account of this third period would be incomplete without mention of other men who have contributed in no small measure to its work—Dr. H. H. Donaldson of Anatomy, and later Dr. C. Judson Herrick, whose course in Neurology has for years been a higher degree requirement, and an opportunity for philosophical speculation as well; and now Drs. Bartelmez and Polyak. Within the laboratory, generations of students have come to know Dr. K. S. Yum, himself a doctor in the Department, who for so many years was Departmental Librarian. Our friends in related Departments who have contributed to the Psychology program are too many to begin to enumerate. Members of this Department have always cooperated with others on border-line or joint-interest problems, and have shared their field of work with psychologists in other Departments. From the first, research and instruction in educational psychology has been the responsibility of the Department of Education, but psychologists in both groups have always been on friendly and cooperative terms, and students have moved back and forth freely in working out their programs. The same kind of relations have been maintained with psychologists in the School of Business, Home Economics (child psychology), Medicine and Psychiatry, and other Schools and Departments having psychological interests; and the work of all these psychologists well deserves more mention than this brief account of the Department permits. In addition, public and private institutions, in city and state, have generously provided facilities for demonstration, research, and apprentice-training without which the Department's work would have been far less effective; we may cite, as representative, the Institute for Juvenile Research, the Chicago Public Schools, and several state hospitals. Psychologists from other Universities have also made their contribution to the Department's program. It has always been the practice to invite visiting professors of reputation, usually during summer quarters. Ladd of Yale and Baldwin of Princeton were early guest-professors. I recall also the names of others, including Scott, Pillsbury, Watson, Sutherland, Joseph Peterson, Faris, June Downey, Yerkes, Warner Brown, Franz, Griffith, Hunter, Stratton, Angier, Bott, Perrin, Rosenow, Herman Adler, Hull, Taylor, K. Buehler, Culler, Paterson, J. J. B. Morgan, Darrow, B. F. Skinner, Grace Munson, and Farnsworth, besides Thurstone, Lashley, and Rogers who later joined the staff, and perhaps others. All of these, therefore, the Department can claim as past members.

V. THE RECENT YEARS

When Mr. Carr retired in 1938, no provision had been made for a successor. Departmental relations with the Biological Sciences Division were a bit disturbed, and for a time the status of the Department was somewhat confused. During the summer and autumn of that year Mr. Thurstone stepped into the breach, and brought into the Department as full time members two men who had been assistant examiners with him while he was University Examiner, and had been serving part time as instructors in the Department, Messrs. Gulliksen and Wolfle; a third, M. W. (now Lt. Col.) Richardson, went to Washington as Chief Examiner in the Civil Service. That winter the Department's administrative affiliation was transferred to the Division of the Social Sciences, retaining, however, representation in the Biological Sciences, and continuing to award the B.S. and Divisional M.S. degrees in that Division. With the transfer came a new Chairman, an educational psychologist of note, Frank N. Freeman, of the Department of Education. At his request the writer assumed the title and duties of Secretary of the Department. Unhappily, Mr. Freeman's chairmanship was very temporary, since he resigned that summer to accept appointment as Dean of the School of Education of the University of California, where he is still active. The Department's administrative responsibilities thereupon devolved upon the Secretary, where they have remained ever since, although the title was superseded in 1945 by that of Acting Chairman.

The war, as we all know, has severely disrupted the Department's work in many ways, not least by the absence for war service of several staff members. These gaps, temporarily filled by J. E. P. Libby, L. R. Tucker, Edith Sherman Jay, Virginia Brown and others, are now provided for on a more enduring basis by the appointment of A. W. Brown, for long a part-time lecturer in the Department, and T. G. Andrews, and for the coming year, that of C. R. Rogers and the expected return of Dr. Wolfle. We look forward hopefully to the appointment of additional staff members, to provide the means of accomplishing our Departmental goal of contributing substantially to the growth and usefulness of psychological science.*

VI. DOCTORATES IN THE DEPARTMENT

But the most important thing in the Department's history is not, of course, its housing and equipment, nor its teaching staff, but its research and training output, including hundreds of graduates who have gone out to continue their careers of productivity and service, and especially those who have received its highest recognition, the Doctorate of Philosophy.

In the 42 years of the Department's independent existence, 160 men and women have earned its Ph.D. degree, 115 men and 45 women. Of these, 39 received the degree during the period of Mr. Angell's chairmanship, 1903 to 1919; 108 between then and 1938, the year of Mr. Carr's retirement; and 13 since that date. Numerically, the high points were 1931 with 12, 1928 and 1930 with 11 each, and 1917 with 10; the low points, years in which no doctorates were awarded, were 1906,

---

* *Postscript, May, 1946.* Some changes since this paper was prepared a year ago may be noted. Dr. Wolfle's resignation to become Executive Secretary of the American Psychological Association was followed by the appointment of William D. Neff as Assistant Professor, and of Virginia Brown, Virginia Axline, and M. H. Groves as Instructors for 1946–47; while negotiations are under way looking to three or more new appointments at professorial levels. The current influx of graduate students, now nearly four times as many as a year ago, has brought the student-load to its peak and intensified instructional, staff, and space problems, so that limitation of admission now seems inevitable.

1918, and 1919 (during the first World War), 1923, and 1944 (the second World War). Even before 1942 Chicago doctors in Psychology were scattered all over the world, including every part of the United States and Canada. To read the titles of the dissertations is of interest, for they give a clear picture of the trend of interests among staff and students of the Department. Any perfect or even reliable classification of these topics is, of course, impossible; but even such a rough classification as the following, which is based primarily on subject-matter, is, I believe, illuminating. The numbers given are percentages of the total output of 160:

Learning, of which 19% were verbal learning, and 13% motor learning	32%
Other motor functions, including emotion and suggestion	9%
Psychobiology, including effects of drugs, diet, vitamin deficiency	9%
Beginning in 1917 with Arlitt's studies of the behavior of alcoholic rats, but mainly done under Lashley	
Individual differences and their measurement	9%
Beginning in 1915 with Kitson's study of college students	
Psychophysics and psychometrics	8%
Plus some classified under other headings	
Personality and social psychology	8%
Throughout the history, beginning with Faris's study on punishment in 1914	
Sensory functions	8%
The majority before 1917	
Perception	6%
Most before 1930	
Other processes—imagery, thought, attention, affection, aesthetics	6%
Most earlier than 1920	
Animal psychology, undifferentiated topics	2%
All before 1920	
Pathological behavior	2%
Child and genetic	1%
	100%

Master's degrees have during this time been awarded to 116 students, 22 of whom later earned their doctorates in the Department and several more in other Departments of the University or elsewhere. I have not attempted to classify their dissertations, but I am sure, from casual inspection, that it does not differ widely from the one just described. We have no count of Bachelor's degrees in Psychology prior to 1932, when the Divisional plan with provision for Departmental programs began; but 268 Bachelors of Arts and Bachelors of Science in Psychology, not quite 21 per year, have been awarded since then.

Until comparatively recent years, most doctoral dissertations (as well as some Masters' and other laboratory studies) found publication in the *Psychological Review Monograph Supplement* (later, *Psychological Monographs*), which Mr. Angell edited for some years. Most of these were parts of a Chicago Laboratory Series. Other studies, especially animal, were published in the *Journal of Comparative Neurology and Psychology*, the *Journal of Comparative Psychology*, and the *Behavior Monographs* (Chicago Series). During the 1920's, the University for some years required publication only of abstracts of theses, in its own series of annual volumes, although during this period some were published in full elsewhere. Since then, as dissertations in psychology have tended to become more concise and less discursive, most have been published in full or in "essential" form, many of them in the *Journal of Experimental Psychology* and *Psychometrika*. But there is hardly a psychological journal in the English language which has not contained one or more studies from the laboratory; while throughout the Department's history, its staff members have served as editors or on the editorial boards of many psychological journals.

This list of 160 doctors includes many names familiar not only to psychologists and educators, but, in some cases, to a much wider public. Most of those later associated with the Department as staff members have been mentioned; while others of the earlier doctors who have achieved recognition through research, publication, administration, or teaching, include many of whom we can name only a few. The first period (Philosophy) includes, among other psychologists, E. S. Ames, D. P. Macmillan (long in the Chicago Schools), Helen Thompson Woolley, and Kate Gordon Moore. The Department's own roster includes such names as Beardsley Ruml, Chi Weh Luh of retention curve fame (the first of 13 Chinese graduates), June Downey, Walter Van Dyke Bingham, Joseph Peterson, C. S. Yoakum, Walter S. Hunter, L. A. Pechstein, Ada H. Arlitt, Margaret Wooster Curti, Helen Koch, E. A. Culler, Dr. William E. Blatz, Thelma Gwinn Thurstone, Chester Darrow, Chaio Tsai, now one of the world's foremost physiologists, the late John McGeoch, Dr. W. H. Sheldon of Somatotype fame, President Goodrich White of Emory University, Dr. Mandel Sherman, Kate Hevner Mueller, Dean of Women at Indiana, former President W. H. Cowley of Hamilton, Marion Monroe, the reading specialist, and dozens of others not less worthy of being named. Quite deliberately I have omitted mention of the 55 holders of Ph.D.'s granted since 1930, for it is still a bit early to appraise fairly their achievements.

Of the 160 doctors, over a hundred have pursued their careers in college and university positions. More than a score of these are or have been heads of University Departments, and a dozen have held major administrative appointments, deanships or presidencies. A cross-section of the psychological and other vocations they represent would probably resemble a fair sample of the membership of the American Psychological Association. Besides psychological teaching and research, Chicago doctors are to be found in many branches of public service, national, state, and municipal (not to mention army, navy, and civilian war work), in educational administration, notably as school psychologists, in industry and business, in counseling and guidance, and in public welfare institutions of many kinds; while at least seven hold medical degrees. I estimate that well over two-thirds of the 160 have actively continued since their doctorates to produce scholarly research and publication, while still others have contributed to the literature of educational, business, or public welfare administration.

Besides our own doctors, seven National Research Council Fellows —C. F. Jacobsen, I. Krechevsky, R. W. Leeper, N. R. F. Maier, L. A. Pennington, T. C. Schneirla, and P. H. Settlage and one General Education Board Fellow, W. A. Varvel, have carried on post-doctoral studies with members of the Department staff, most of them with Lashley; while through the years, a large number of other visiting doctors have worked in the Department's laboratories for longer or shorter periods.

VII. THE DEPARTMENT AND PSYCHOLOGICAL ORGANIZATIONS

It is of interest to note that the Psychological Journal Club is even older than the Department. A Philosophy Club first announced in the 1894–95 Register was supplemented in 1901–02 by a Psychological Journal Club, meeting bi-weekly, its principal work being to present critical reports upon current experimental literature. This function continued to characterize the club's activities until about 15 to 18 years ago—I do not recall the exact date. At that time, long-continued efforts to bring undergraduates and first-year graduate students into more active participation led to the formation of a Junior Psychological Club, whose weekly meetings featured an address, often by a visiting lecturer from some other institution or Department. Meanwhile, the advanced students continued the Senior Journal Club along established lines. But the Junior Club became so popular that it attracted the members of the older club, and after a year or two the Senior Club was discontinued, and the Psychological Club continued to function after the new pattern. Most of us have felt that the change brought a real loss to the Department's training function; and while current dissertations and other researches have been presented, only within the last year or two has the earlier "Journal" function been revived. Many of us hope that this valuable opportunity for public appearances by embryo psychologists may resume something of the importance it once had.

Psi Chi, the honorary psychological society founded in 1929, has from the start had a chapter here, but it has seldom played the important role it has in many institutions, because of the older and stronger tradition of the Journal Club. However, the current movement to revive and increase its functions and membership promises well for its continuing usefulness at Chicago. Meanwhile, the national organization has never been without important Chicago contacts; the Department's alumni have furnished two national presidents, McKinney and Fritz, each serving two years, and two national secretaries, Louise Grossnickle Yum and Dorothea Ewers. Another national honorary organization, the scientific society of Sigma Xi, has over many years elected to its membership a good many prospective doctors in Psychology; while the Department's bachelors have always been represented among those elected to Phi Beta Kappa.

The Department has also had recognition by national and regional associations. The American Psychological Association has on five occasions been the guest of the Department, in 1901, 1907, 1915, 1920, and 1933; and nine of its past presidents, the highest honor bestowed by American psychologists, have been Chicago doctors or staff members—Dewey, Angell, Judd, Watson, Carr, Lashley, Hunter, Thurstone, and Peterson, as well as several others who have had summer-quarter visiting appointments in the Department. The Midwestern Psychological Association has met here twice since its establishment in 1926, and its presidents have included five Chicago alumni, Thurstone, McGeoch, Carr, Bills, and Culler. So the Department has played its role not only locally but regionally and nationally.

Members and alumni of the Department have reason to look back with something of pride upon its record of accomplishment. But a worthy past is not enough. We look forward with hope and confidence that the years to come will bring a still more noteworthy record of Departmental contribution to the growth and usefulness of our science.

NEW STATISTICAL CRITERIA FOR LEARNING AND PROBLEM SOLUTION IN EXPERIMENTS INVOLVING REPEATED TRIALS

DAVID A. GRANT
University of Wisconsin

Psychologists generally have barely begun to utilize the variety of statistical techniques available to demonstrate learning and problem solution in experiments involving repeated trials. Many studies still appear using arbitrary criteria unsupported by probability statements, in spite of their availability and desirability. Also some methods have become available only recently and for that reason have not been used. The present paper includes a brief discussion of previously used procedures and presents two new criteria which have not been used in psychological research, but which seem to offer possibilities for more complete analyses of learning, problem solution, and other forms of behavior which have been studied by means of repeated trials. Some of the types of experiments to which these techniques are applicable include simple and complex discrimination experiments, variations of the Hamilton "perseverance" problem, food and other preference tests, spatial and mental mazes, string-pulling tests, Weigl and other sorting tests, delayed reaction experiments, conflict experiments and simple social reaction studies.

PREVIOUSLY USED CRITERIA

All experiments involving repeated trials on the same basic task give rise to problems of assessing degree of improvement. In most instances the subject may "solve" the problem by chance, and a certain proportion of his correct responses are therefore suspect. The exact number of chance solutions cannot, however, be determined, and the experimenter must use some indirect procedure to decide whether or not his subject's performance is other than what might be "expected" by chance. There have been two frequently used ways of determining whether a problem has been learned or solved. A non-statistical arbitrary level of performance has been the most widely used criterion even in the recent past. For example, a rat is required to run the maze with no errors one or more times (8), or the subject must select the correct one of two stimulus objects 18 out of 20 times, getting the last ten correct (9). The other frequently used criterion is set up in terms of the probability of the subject's making a certain *total number of correct responses* by chance (5, 7, 16). With this method, the experimenter assumes that the subject has learned the task if the probability of his responding correctly by chance as many times as he did actually respond falls below a conventional (arbitrary) level.

The use of total correct responses, whether interpreted in terms of arbitrary or statistical norms, has serious limitations when an attempt is being made to investigate anything beyond the most restricted aspects of learning or problem solving behavior. About the only thing to be said in favor of the arbitrary level without statistical interpretation is that it is objective. Statistical criteria based upon total correct responses are satisfactory only if the experimenter is concerned with overall performance. Total correct responses are only one aspect of performance, and other aspects need to be considered in analytic work. This is well-illustrated by some data of Young on food preference in the rat (**16**, p. 360). Wheat and sugar were presented simultaneously for 105 trials; selection of wheat is indicated by *W;* selection of sugar is indicated by *S*. The 105 selections in order were: *WWWWWSSSSSSSSSSSSSSSSSSWWWWWWWWWWWWWWWWWWWWWWWWWSWWWWWWWWWWWWWWWSSSSSSSSSSSSSSSSWSSSWWWWWSWWWWWWWWWWWWWSS*. Young remarks very correctly that, "A mere count of relative frequency does not adequately portray what is happening." The *total correct* statistical criterion as adapted for these data gives no proof or indication of the statistical significance of the striking "runs" of selections made by this subject.

The new techniques to be described in the present paper are statistical tests which should permit the experimenter to demonstrate "what is happening" more completely than he has heretofore been able to do.

Interpretation of Statistical Criteria

The reasoning behind the statistical tests to be described follows a familiar pattern. Consider an experiment set up so that the subject is given n trials with a constant probability,* p, of responding correctly. Krechevsky's "Hypotheses" experiment (**7**) serves as a good example. In it, the night's work for each subject consisted of solving 40 discriminations—*i.e.*, $n=40$—each of which involved a choice of one of *two* brightnesses—thus $p=.5$. To demonstrate that the subject has learned or solved the problem, it is first necessary to hypothesize that he is responding at random with respect to the critical variable;

---

* Using the classical definition, the probability of the occurrence of a correct response is obtained by enumerating the total number of equally likely, mutually exclusive possible responses under the specified conditions and dividing this number into the number of equally likely, mutually exclusive responses designated as correct.

i.e., that he has not learned.† Next is deduced mathematically the probability of a chance performance equal to that of the subject. Sometimes this is done directly, but more often it is convenient to compute the expected (average) random performance and some measure of series to series variability from the expected value. Then the subject's actual performance is considered as a chance deviation from the expected value in terms of the measure of variability, and the mathematical probability of such a chance deviation is assessed. Then, finally, if the probability of a chance deviation as great or greater than that represented by the actual performance falls below a certain level the experimenter will reject the hypothesis that the behavior is random with respect to the critical variable. This constitutes the statistical test. The experimenter then assumes that inasmuch as the critical variable was randomized with respect to the other variables his subject has learned or solved the problem.

THE CRITERION OF TOTAL NUMBER OR PROPORTION OF CORRECT RESPONSES

Before considering new criteria, the commonly used statistical criterion of total number or proportion of correct responses will be discussed. The mathematics of this test involves determining the probability that the actual number of correct responses equals or exceeds a given number, m, in a sequence of n independent trials where the probability of a chance success on a single trial is p. This involves the Bernoulli distribution, which is dealt with by most books on mathematical probability or statistics (**1**, p. 197; **13**, p. 44). The exact probability of m correct responses can be shown to be:

$$P_m = C_m{}^n p^m q^{n-m}, \tag{1}$$

where $C_m{}^n = \dfrac{n!}{m!(n-m)!}$, $n!$, read n-factorial, equals $1 \cdot 2 \cdot 3 \cdots n$, and $q = 1 - p$. The experimenter is interested in the total probability that his subject might by chance do as well *or better* than m successes, and this quantity will, of course, be equal to the sum of the probabilities of $m, m+1, m+2, m+3, \cdots, n$ correct responses. Written as a formula, the total probability that the number of correct responses equals or exceeds m is:

---

† Note that this does not necessarily assume that the subject is responding entirely at random within the experimental situation. He may be responding by position, by simple alternation of position, or show some other "hypothesis" behavior, but the reasoning does assume that the critical variable, whatever it is, has been varied randomly with respect to all other discriminable stimuli.

$$P(m) = \sum_{i=m}^{n} C_i^n p^i q^{n-i}.^{*} \qquad [2]$$

As a concrete example of the use of [2],† suppose that a rat responds correctly in 19 out of 25 trials in an experiment such as that of Krechevsky, mentioned above. In this instance $p = .5$, $n = 25$, and it is necessary to find the total probability, P, that the performance equals or exceeds an m of 19. Using [2], $P(m \geqq 19)$ will be:

$$\frac{25!}{19!6!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{20!5!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{21!4!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{22!3!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{23!2!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{24!1!}\left(\frac{1}{2}\right)^{25} + \frac{25!}{25!0!}\left(\frac{1}{2}\right)^{25}.^{**}$$

Which, accurate to eight places (in order to use all terms for illustrative purposes), is: $.00527799 + .00158339 + .000377001 + .00006855 + .00000894 + .00000075 + .00000003 = .00731665$. This would be rounded off to .0073. It means that the probability of the subject's getting 19 or more correct by chance out of 25 trials would be .0073, or that about 73 times in ten thousand such 25-trial sequences would the subject be expected to do as well as or better than it actually did if it were responding at random with respect to the critical variable. In this instance the experimenter could reject the hypothesis of random responses at the 1% confidence level.

Formula [2] is not particularly complicated and should be used wherever applicable, but in spite of this, most such total probabilities have in the past been determined by means of the normal curve which gives a rough approximation to the true value of $P(m)$. To enter the table of the normal curve, the normal deviate, x/σ, must be computed (*cf.* 4, Table 15, p. 110). In this fraction $x = m - np$, and $\sigma = \sqrt{npq}$; thus:

$$\frac{x}{\sigma} = \frac{m - np}{\sqrt{npq}}. \qquad [3]$$

---

* This is the familiar formula for Total Probability (**1**, **13**). The index, i, in this notation shows the part of the formula which varies in forming each successive term in the sum. The limits of summation below and above the summation sign indicate that in the first term $i = m$, in the next term $i = m + 1$, etc., until finally i equals n in the last term (*cf.*, **15**, Ch. 6).

† Square brackets will be used in the present paper in referring to the various formulae.

** $0! = 1$.

To find the approximate probability of 19 or more correct out of 25 trials, as in the example above, using formula [3], and the table of the normal curve, $\frac{x}{\sigma}=\frac{19-12.5}{\sqrt{25\cdot(.5)(.5)}}$, a normal deviate of 2.6. This would be interpreted from a table of the normal curve to give a probability of about .0047. The very considerable error of the approximation is thus .0047 − .0073, or −.0026.

The probability obtained from [2] may also be approximated by use of the χ^2 distribution for one *degree of freedom*. With this method the number of correct responses may be designated m, and the number of incorrect responses may be designated s; then χ^2 is given by the equation (2, p. 334):

$$\chi^2 = \frac{\left(m - \frac{p}{q}s\right)^2}{\frac{p}{q}n}, \qquad [4]$$

which for $p=q=.5$ reduces to:

$$\chi^2 = \frac{(m-s)^2}{n} = \frac{(m-s)^2}{m+s}. \qquad [5]$$

In the above example $\chi^2=\frac{13^2}{25}=6.76$. The probability is found from tables of χ^2 to be less than .01 but greater than .005 (**3**, **12**).

In general, equation [2] should be used to compute the probabilities for the Bernoulli distribution, but the normal curve and χ^2 may be used under certain conditions. The limiting conditions for the use of [3], [4], and [5] are as follows: for [3] the normal curve gives very inaccurate approximations if p or q are nearly unity or if n is not fairly large ($n>30$); for use of [4] and [5] both np and nq must exceed five, which means that n must in all cases exceed 10, and if p and q do not equal .5, n must be greater than 10. In general [4] and [5] will give more accurate approximations than [3] when n is small.

Runs as a Criterion of Learning and Problem Solution

The responses of a successful subject are characterized by features other than a high *proportion* of correct responses, and some of these other features may be used as criteria of solution. Suppose that a subject responds randomly to the two-choice discrimination problem for the first fifteen trials, and then responds correctly for the last ten trials. The total correct criterion would fail to demonstrate learning for this subject. It is, however, possible to determine the total probability of at least one "run" of length s or greater in a sequence of n trials with a constant probability of p of a chance success on any one trial. The probability of a run of ten in a series of 25 with a single-trial probability of .5, as in the data above, would be .0083, and an experimenter, if he accepted data at the conventional 1% confidence level as significant, would assume that his subject had solved the problem.

In general, the probability that there will be one or more runs of s or more chance successes in n trials where the probability of a chance success on each trial is p will be: $P(s) =$

$$1 - \sum_{i=0}^{n/(s+1)} (-1)^i C_i^{n-is} (qp^s)^i + p^s \left[\sum_{j=0}^{(n-s)/(s+1)} (-1)^j C_j^{n-s(j+1)} (qp^s)^j \right]. \quad [6]$$

Equation [6] is based upon the solution given by Uspensky (**13**, p. 74) As an example of computation, using [6], consider the probability of one or more runs of ten or more in a series of 25 trials as in the paragraph above. In this case [6] becomes:

$$1 - \sum_{i=0}^{2} (-1)^i C_i^{25-10i} \left(\frac{1}{2}\right)^{11i} + \left(\frac{1}{2}\right)^{10} \left[\sum_{j=0}^{1} (-1)^j C_j^{25-10(j+1)} \left(\frac{1}{2}\right)^{11j} \right]$$

$$= 1 - 1 + C_1^{15} \left(\frac{1}{2}\right)^{11} - C_2^{15} \left(\frac{1}{2}\right)^{22} + \left(\frac{1}{2}\right)^{10} \left[1 - C_1^5 \left(\frac{1}{2}\right)^{11} \right]$$

$$= 1 - 1 + \frac{15}{2048} - \frac{10}{(2048)^2} + \frac{1}{1024}\left(1 - \frac{5}{2048}\right) = .0083.$$

Computations with [6] are fairly elementary but rather tedious, and therefore the probabilities for various length runs with $p = .5$ are given for convenient n's in Table 1. The use of Table 1 is extremely simple. By selecting the column corresponding to the number of trials in the sequence and following it down, one can find the probability of a run of any length from six up to runs which are significant at the 0.1% confidence level. Evaluation of probabilities from the table may be illustrated with some unpublished discrimination data obtained by Harlow in the Wisconsin Primate Laboratory from a monkey solving a Spence-type (**9**) problem.* Correct choices are indicated by R, and incorrect choices are indicated by *W*. The data are as follows: *WRWRWWRWWRRRRRRRRRWR*. The longest run is nine correct choices in the series of 20 trials. The probability of a run of nine in 20 trials is given in Table 1 as .0127, which means that such data would be expected by chance less than twice in a hundred such 20-trial sequences. Note that the data do not approach significance in terms of the total correct criterion.

---

* In this problem the subject had to select a stimulus object which has been incorrect in previous discriminations.

TABLE I

PROBABILITIES OF ONE OR MORE RUNS OF S OR MORE CHANCE SUCCESSES IN A SERIES OF N TRIALS WITH A CONSTANT SINGLE-TRIAL PROBABILITY OF 0.50

S\N	16	20	25	50
6	.0929	.1223	.1577	.3146
7	.0429	.0582	.0770	.1653
8	.0195	.0273	.0369	.0836
9	.0088	.0127	.0176	.0415
10	.0039	.0059	.0083	.0204
11	.0017	.0027	.0039	.0100 +*
12	.0007	.0012	.0018	.0049
13	—	.0006	.0009	.0024
14	—	—	—	.0018
15	—	—	—	.0008

* This value is slightly over .0100, so that a run of 12 would, therefore, be considered significant at the 1% confidence level.

The criterion of runs will prove most useful in instances where the mode of response is likely to shift suddenly—so-called insightful or hypothesis behavior—and when little oscillation (6) may be expected to occur. In data where oscillation is appreciable, the total correct or some other criterion should be used. A variation of the runs criterion may prove to be very useful in cases of partial learning where there is considerable oscillation and also in analytic study of the learning process. This variation is described below.

THE GROUPING OF SUCCESSES AND FAILURES AS A CRITERION OF LEARNING AND PROBLEM SOLUTION

During a sequence of random responses, the subject tends to shift frequently from one response to another in the same fashion that heads and tails alternate frequently in a random series of coin tosses. An example of such a random series of ten coin tosses is:

$$T, H, T, T, H, T, H, H, H, T.$$

The series consists of four groups of one or more *tails*, and three groups of one or more *heads*, making a total of seven groups. With five heads and five tails—as there are here—the maximum possible number of groups would be ten, and the minimum possible would be two. In general, the greater the systematization or non-randomness of a behavior sequence, the fewer* will be the number of groups. A test of randomness may, therefore, be made in terms of the probability of the actual number of groups, u, in the data being less than or equal to any given number u'.

If m is the number of responses of one sort and n is the number of responses of the other sort ($m \leqq n$), then Stevens (**10**) and others (**14**) have found the probability that u is less than or equal to u' to be:

$$P(u \leqq u') = \frac{m!n!}{(m+n)!} \sum_{u=2}^{u'} f_u, \qquad [7]$$

where

$$f_u = 2C_{k-1}^{m-1}C_{k-1}^{n-1} \text{ if } u = 2k \text{ (i.e., } u \text{ is even)},$$

and

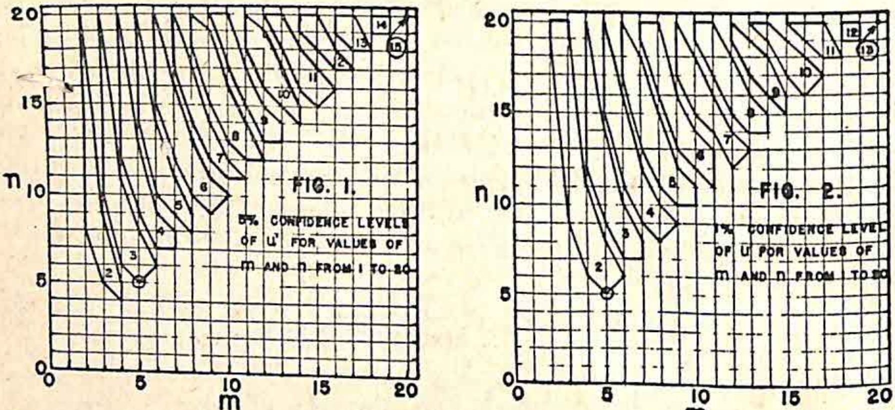
$$f_u = C_{k-1}^{m-1}C_{k-2}^{n-1} + C_{k-2}^{m-1}C_{k-1}^{n-1}, \text{ if } u = 2k - 1, (\textit{i.e.}, u \text{ is odd}).$$

Tables of this probability function have been published by Swed and Eisenhart (**11**). For convenience the 5% and 1% confidence levels of u' have been graphed from Swed and Eisenhart's Table II in Figs. 1 and 2, respectively, for values of m and n up to 20.

Use of Figs. 1 and 2 may be illustrated from the coin tossing example given above. Here m and n both equal 5, and this point has been encircled in each figure. In Fig. 1 the encircled point lies in the area for $u'=3$. Hence it may be concluded that the example, in which there are seven groups; *i.e.*, $u=7$, does not depart significantly from chance grouping. (In Fig. 2 the point corresponding to $m=n=5$ lies in the area for $u'=2$. Significance at the 1% level, of course, requires an even smaller number of groups than significance at the 5% level.) As another example of the use of Figs. 1 and 2, consider the following raw data from a "problem-finding" study on the rhesus monkey (**17**): *RRWWWWWWRRRRRWRRRRRRWWWRR*. Here there are ten wrong responses and 15 right responses, forming *seven* groups. From Fig. 2 it may be seen that when m is ten, and n is fifteen, seven groups will be significant at the 1% level, hence the above series might be said to show systematization or partial learning. Note that neither the total correct nor the runs criterion would show the significant departure from a random sequence.

---

* Systematization may also be shown by an unusually large number of groups. This will not be as generally useful to psychologists as the test described. Consult original references.



FIGURES 1 AND 2. THE 5% AND 1% CONFIDENCE LEVELS FOR u', PLOTTED FOR VALUES OF m AND n FROM ONE TO 20.

In using these Figures, the point determined by the values of m and n ($m \leqq n$) is located in each Figure. This point will, in each Figure, lie in one of the irregularly shaped numbered areas. The number of the area is the 5% or 1% value of u' in Figures 1 and 2, respectively.

The grouping test differs from the other tests discussed in the present paper in that no defined p, or single-trial probability need be specified. Thus non-random grouping of wins or losses in an experiment on competition could be demonstrated, or a sequence of social contacts between nursery school children might be investigated for grouping of approach or avoidance responses, or the positive and negative discrepancy scores of a single subject in a level of aspiration experiment might be analyzed. This generality makes the grouping test far-reaching in its applications. The only assumption involved is that the variable behind the measures is continuous in the general population, and the null hypothesis is that the items are selected independently at random; *i.e.*, all grouping is random.

The grouping criterion is usually most sensitive when m and n are about equal—a situation when the other criteria usually fail to demonstrate non-random sequences. It is therefore likely that this criterion may be of especial use in the stages of learning and problem solution when partial and temporary solutions introduce the first systematization into the behavior sequences.

Another instance in which the grouping (and also the runs) criterion are especially useful would be in making tests of incorrect modes of solution such as Krechevsky made (7), using the total correct criterion. Some subjects; *e.g.*, human adults, shift away from incorrect modes of response so quickly that it is difficult to "catch" them with a statistical test. The abrupt shifts however, are advantageous with the *grouping* criterion and the *runs* criterion. It is not feasible to list all possible uses to which these tests may be put, but it is hoped that there will be a large number of experimental situations in which these new tools will be found to be helpful. Future research and analysis in the various areas in psychology will show the extent of their usefulness, and the interpretations to be attached to results obtained by their application.

SUMMARY

The logic of statistical tests of learning and problem solution in repeated trials was outlined. Statistical tests already in use based upon *total correct responses* in a sequence and new statistical tests in terms of (a) *length of runs of correct* (*or incorrect*) *responses*, and (b) *grouping of successes and failures* were described and illustrative examples were given along with tables and figures to facilitate the use of the new tests. Suggestions were given as to possible applications of the new tests.

BIBLIOGRAPHY

1. Camp, B. H. *Mathematical part of elementary statistics.* New York: Heath, 1934.
2. Croxton, F. E. & Cowden, D. J. *Applied general statistics.* Prentice-Hall, 1943.
3. Fisher, R. A. & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver and Boyd, 1938.
4. Garrett, H. E. *Statistics in psychology and education.* 2nd Ed. New York: Longmans, Green, 1938.
5. Harlow, H. F. & Young, M. L. Generalization by rhesus monkeys of a problem involving the Weigl principle using the oddity method. *J. comp. Psychol.*, 1943, **36**, 201–216.
6. Hull, C. L. *Principles of behavior.* New York: Appleton Century, 1943.
7. Krechevsky, I. "Hypotheses" versus "chance" in the pre-solution period in sensory discrimination-learning. *Univ. Cal. Publ. Psychol.*, 1932, **6**, 27–44.
8. Lashley, K. S. Sensory control and rate of learning in the maze. *J. genet. Psychol.*, 1945, **66**, 143–145.
9. Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzees. *J. comp. Psychol.*, 1937, **23**, 77–100.
10. Stevens, W. L. Distributions of

groups in a sequence of alternatives. *Ann. Eugen.*, 1939, **9**, 10–17.

11. Swed, Frieda S. & Eisenhart, C. Tables for testing the randomness of grouping in a sequence of alternatives. *Ann. Math. Statist.*, 1943, **14**, 66–87.
12. Thompson, Catherine M. Table of the percentage points of the χ^2-distribution. *Biometrika*, 1941, **32**, 187–191.
13. Uspensky, J. V. *Introduction to mathematical probability.* New York: McGraw-Hill, 1937.
14. Wald, A. & Wolfowitz, J. On a test whether two samples are from the same population. *Ann. Math. Statist.*, 1940, **11**, 147–162.
15. Walker, H. M. *Elementary statistical methods.* New York: Holt, 1943.
16. Young, P. T. Studies of food preference, appetite and dietary habit: I. Running activity and dietary habit of the rat in relation to food preference. *J. comp. Psychol.*, 1944, **37**, 327–370.
17. Zable, Myra & Harlow, H. F. The performance of *rhesus* monkeys on series of object-quality and positional discriminations and discrimination reversals. *J. comp. Psychol.* (In press).

DeJong, H. H. *Experimental catatonia.* Baltimore: Williams and Wilkins, 1945. Pp. xiv+225.

DeJong defines and describes human catatonia in terms of the hypokinetic and hyperkinetic psycho-motor symptoms. Among the hypokinetic phenomena are diminished motor activity, catalepsy, and negativism. Among the hyperkinetic phenomena are impulsiveness, aggressiveness, and symptoms which depend upon activity of the autonomic nervous system, such as excessive salivation and blood volume changes. In experimental catatonia in animals the hypokinetic phenomena are shown by diminished motor activity, catalepsy, and physiological negativism. The hyperkinetic phenomena are shown by running, jumping, and by autonomic nervous system symptoms.

Having thus defined experimental catatonia and its relation to the catatonia seen in psychiatric patients, DeJong describes and gives protocols for a large number of experiments done by himself and his collaborators in the production of experimental catatonia in a variety of animal forms. His original experiments were done with bulbocapnine. The injection of this drug brings about a decreased motor activity, negativism, catalepsy, and salivation, which seem quite similar to human catatonia. He demonstrated that the action of the drug is directly upon the central nervous system. A series of experiments indicated that the drug did not act on any particular portion of the central nervous system, since it was impossible to interfere with the catatonic symptoms by the ablation of any particular part of the brain.

DeJong worked through a long series of other drugs, trying to find which ones would bring about the catatonic symptoms as well as the determination of the particular molecular form or portion of the chemical which might be active. He found a long list of chemical substances which would bring about catatonic symptoms, but he was not able to solve the problem of whether or not there existed a chemical "catatonizing nucleus."

The work with the chemicals indicated that symptoms were due to an interference with the oxygen supply or oxygen utilization by the brain. On the basis of this hypothesis he conducted a series of experiments showing that partial asphyxiation brought about in a variety of ways did produce the catatonic syndrome. Whether or not this was due to oxygen lack was not finally demonstrated since asphyxiation is accompanied by failure in other nutritional supplies and by a change in the acid-base ratio of body fluids.

Another series of experiments was carried out attempting to produce catatonic symptoms by neuro-surgical interference with various brain structures. No regular or conclusive results were obtained. He found that rats placed in a centrifuge and spun rapidly showed catatonic symptoms. In all probability the symptoms were due to brain injury and to the interference with brain metabolism.

He next reports a series of experiments carried out with rats, in which he produced catatonic symptoms by electrical stimulation and by loud, continuous sound stimuli. He was unable to produce catatonic symptoms in humans when similar stimulating conditions were employed. A series of experiments were conducted in which catatonic phenomena were produced in dogs and cats by alterations in metabolic function brought about by surgical interference with the orderly function of the intestine and of the liver. He interprets these changes as due to some toxic substance produced or released by the liver as a result of the operation.

Experiments were conducted with human patients in an attempt to isolate possible toxic substances in the urine of the catatonic patient. He found that a substance which was closely allied to nicotine did produce catatonic symptoms in animals. However, it was not specific to catatonic patients, being recovered from the urine of certain normal humans. He then conducted a long series of experiments working through groups of chemicals more or less allied to nicotine and to histamine. He was unable to show that any one of these substances appeared in greater concentration in the blood of the catatonic patient than it did in his normal control group.

He sums up the methods by which experimental catatonia may be produced as follows:

a. *chemical:* bulbocapnine, cannabis indica, mescaline, substances preceding mescaline-synthesis, certain gases, substances derived from adrenaline, mescaline and choline, nicotine;

b. *biochemical:* epinephrine, acetylcholine, asphyxiation, substances from human urine, benzoilic urinary extracts;

c. *alteration of metabolic function:* ligation of the hepatic artery, ligations of the small intestine, Eck's fistula;

d. *electrical:* Leduc's current, faradic current, electroshock;

e. *neuro-surgical:* various brain-lesions;

f. *miscellaneous:* bilateral ligation of the carotid artery, centrifugation, audiogenic stimuli, "animal hypnosis" (only hypokinetic signs).

In light of this evidence he concludes that biologically human catatonia is in all probability closely related to some alteration of the function of the liver and the intestine which results in some toxic action on the brain cortex. This conclusion is in close agreement to the hypothesis advanced over fifty years ago by Kraepelin.

This is a report of a well-defined, scientific approach, in which the problem is stated and the resulting hypothetical deductions are experimentally tested. Not all of the experiments are adequately described, nor are all of them conclusive. However, the bulk of the work, together with the unity of the evidence, is impressive. He has been able to show a great degree of similarity between human catatonia and experimental catatonia. He has been unable to specify the basic biological or physiological mechanism which produces the catatonic symptoms. He is to be commended for the thoroughness of his experimental approach. One might venture a prediction that the solution of the problem of schizophrenia will be found by the systematic continuation of the experimental approach which DeJong has so conscientiously followed.

CARNEY LANDIS.

Psychiatric Institute,
Columbia University.

HARROWER, ERICKSON, M. R. & STEINER, M. E. *Large scale Rorschach techniques. A manual for the Group Rorschach and Multiple Choice Test.* Springfield, Ill.: Charles C Thomas, 1945. Pp. xi+419.

The late Max Wertheimer once described the relationship between his systematic position and that of Kohler and Lewin as follows: "Kohler is my son and Lewin is my nephew." A similar relationship may be said to exist between the individual Rorschach Method and the two techniques dealt with in this book. *The Group Rorschach* is a son whose claim to legitimacy is still in doubt, and the *Multiple Choice Test* is a second cousin. That it is necessary to make clear the distinction between the techniques is evident from the fact that in a recent standard textbook on abnormal psychology, the Group Rorschach and the Multiple Choice Test are confused with each other. Furthermore, the results from a study using the latter apparently are employed in drawing a conclusion as to the Rorschach Method.

This book purports to be a manual for the Group Rorschach (a technique of administering the Rorschach ink-blots to groups) and for the Multiple Choice Test (a technique in which the subject selects his responses to the Rorschach ink-blots from a list). It also presents certain normative material as well as the results of studies dealing with specific aspects of both techniques.

The material on the Group Rorschach is frankly designed for the use of persons skilled in the Rorschach Method. The literature on the Group Rorschach is well summarized, and a convincing case for the usefulness and merit of this approach is made. The authors contend on the basis of one study and a passing reference to two others that the group method yields essentially the same material as that yielded by the individual Rorschach. According to Dr. Marguerite Hertz, however, other studies not cited by the authors do not bear out this contention.*

---

* This point is made by Dr. Hertz in her review of the book in the *Rorschach Research Exchange* for March 1945. The entire review is recommended for those with a technical interest in the Rorschach Method.

In the section serving as a manual for the Group Rorschach, excellent directions for administration are given. The recommended methods for conducting the inquiry for determinants involve, however, what is tantamount to the use of leading questions—a practice which, for the individual Rorschach, both Beck and Klopfer strongly condemn. The chapter on scoring is not intended to aid the prospective scorer of the Group Rorschach since the authors believe that the scoring principles of the individual Rorschach apply. It seems likely, however, that all "F-" scores might not be the same with the two techniques and one might infer even from the authors' own data that "popular" responses and probably "usual detail" responses do not correspond entirely with their counterparts in the individual Rorschach.

Over 200 pages are devoted to data concerning locations, determinants, and content—mostly in graphic and tabular form. Unfortunately the value of this material is seriously weakened by the following major and minor flaws.

1. No account of the procedure of administration of scoring is given.
2. The sampling is inadequate. Of the 340 subjects tested, 224 were of college age, and the rest consisted of relatively small groups of adult males, prisoners, and a mixed group of psychotics and psychopathic personalities. It is very doubtful whether any meaningful differences can be found between groups so heterogeneous within themselves.
3. No measures of dispersion or measures of reliability of differences are presented.
4. The authors state that their "main interest and concern [was in] providing norms for the college age group" (4), but since their group is composed mainly of medical students and nurses, even this aim is not fulfilled.
5. Instead of attempting to ascertain what are the "usual details" from their own data, the authors use Klopfer's series and add other details which were seen by their subjects at least as often as the most infrequent one he lists. This results in their identifying as *usual* a large number of details whose actual frequencies are less than 1%, i.e., *unusual* "usual details."
6. Although average figures for the different determinants would have been helpful in interpreting a Rorschach record, only percentages are given.

On the jacket flap, the following description of the Multiple Choice Test is offered: " . . . by utilizing their research derived from large scale investigations they have devised an entirely new test which can be readily used for screening out maladjusted persons by investigation *without extensive training in the Rorschach Method*." Although the original validating study of the authors showed a very good discrimination between normals and abnormals, subsequent studies (one of which is presented in full in the book) have found such imperfect discrimination as to make its value as a screening instrument very dubious. The authors attempt to explain away these findings, but the more recent study of Jensen and Rotter showing that the test places 45% of "actually excellent [Army] officers" in the abnormal group makes such attempts rather futile.

In an effort to overcome some of the shortcomings of the test, the authors developed an amplified form. In this version, the number of items was tripled, objectivity of scoring was increased, differential weighting was introduced and a new critical score adopted. These changes were made on the basis of an analysis of "more than 50,000 test blanks," but only fragmentary data are presented to justify most of the innovations. The reliability and validity of the new instrument are not mentioned. Fortunately a recent study allows a tentative evaluation of this form. R. F. and D. I. Malamud, comparing the responses of 100 normals with those of 100 abnormals, found an almost complete overlapping of the two distributions. Item analysis, with the standard of C.R. or 2 or over, revealed only 10% of the items as discriminative while 6% discriminated in the wrong direction.

Some material of value to those interested in experimental work can be gleaned from the book, but these hardly compensate for the many minor inaccuracies, the unsupported speculations, the overly optimistic attitude of the authors, the lack of organization of the material and the unduly low scientific standards employed. Since it is clear that both the Group Rorschach and the Multiple Choice Test are still in the experimental stage, the book offers little to the clinician.

ROBERT C. CHALLMAN.

Psychological Laboratories,
Norwich State Hospital, Norwich, Conn.

ROUSSEAU, JACQUES. *L'heredité et l'homme.* Montreal: l'Arbre, 1945. Pp. 250.

This volume is published by France Forever, a Canadian affiliate of the French Committee for National Liberation, as part of an effort to keep French science alive. The author, who has written three previous books on genetics, appears to have aimed this volume for popular consumption, writing in simple witty style, and ranging from descriptions of the mechanisms of heredity to how to bring up children.

When the author writes about genetics no fault can be found, but when he crosses to psychology he appears to be less scientific and more vulnerable. He believes that nomadism and alcoholism are inherited factors, states that the existence of the criminally-born is definitely proven and after discussing the nature-nurture controversy states: "It is natural to conclude that the IQ is fixed by heredity."

A few incautious statements mar an otherwise excellent book. Altho an adequate glossary and index are provided, the bibliography is fragmentary and the author annoyingly does not further identify people whom he quotes.

RAYMOND CORSINI.

Reception Center,
Elmira, New York.

BOOKS AND MATERIALS RECEIVED

ALEXANDER, F. & FRENCH, T. M. *Psychoanalytic personality. Principles and application.* New York: Ronald Press, 1946. pp. xiii +353.

BLAIR, G. M. *Diagnostic and remedial teaching in secondary schools.* New York: Macmillan, 1946. Pp. xv+422.

BRODY, SAMUEL. *Bioenergetics and growth. With special reference to the efficiency complex in domestic animals.* New York: Reinhold Publishing Corp., 1945. Pp. xii+1023.

DEUTSCH, A. *The mentally ill in America.* New York: Columbia Univ. Press, 1946. Pp. xvii+530.

GUTHRIE, E. R. & HORTON, G. P. *Cats in a puzzle box.* New York: Rinehart & Co., 1946. Pp. 67..

HARRIMAN, P. L. (Ed.) *Twentieth century psychology.* New York: Philosophical Library, 1946. Pp. xiii+712.

MORGAN, J. J. B. *How to keep a sound mind* (Rev. Ed.) New York: Macmillan, 1946. Pp. vii+404.

ROGERS, CARL & WALLEN, JOHN L. *Counseling with returned servicemen.* New York: McGraw-Hill Book Co., 1946. Pp. vii+159.

SEWARD, GEORGENE H. *Sex and the social order.* New York: McGraw-Hill Book Co., 1946. Pp. xi+301.

SMITH, B. L., LASSWELL, H. D. & CASEY, R. D. *Propaganda communication and public opinion.* Princeton: Princeton Univ. Press, 1946. Pp. viii+435.

The march of medicine. New York Academy of Medicine Lectures to the Laity No. X. New York: Columbia Univ. Press, 1945. Pp. xiv +154.

1945 fall testing program in independent schools and supplementary studies. Educational Records Bulletin No. 44. New York: Educational Records Bureau, 437 W. 59th St., January, 1946.

Theory and practice in historical study: a report of the committee on historiography. Social Science Research Council Bull. 54. New York: Social Science Research Council, 1946. Pp. xi+177.

Vol. 43, No. 4 July, 1946

Psychological Bulletin

OPINION-ATTITUDE METHODOLOGY

QUINN MCNEMAR*

Stanford University

I—Problems and Issues. II—Attitudes by Scaling Techniques. III—Single Question Opinion Gauging. IV—Administration. V—Statistical Issues. VI—Study of Changes. VII—Correlates and Interrelationships. VIII—Studies of Morale. IX—Concluding Remarks.

I. THE PROBLEMS AND ISSUES

Over-all View. This paper presents an appraisal, by an outsider, of the techniques and methodologies utilized in studies of opinions and attitudes. First, a brief word as to the meaning of the terms *opinion* and *attitude*. The common element of most definitions of social attitude is that such an attitude is a readiness or tendency to act or react in a certain manner. No one has ever seen an attitude; an attitude, however real to its possessor, is an abstraction the existence of which is inferred either from nonverbal overt behavior, or from verbal or symbolic behavior. The term opinion is frequently defined as the verbal expression of an attitude. This could mean that one can never hold an opinion unless it is expressed, and that in "gauging public opinion" one goes out to gauge something which may be nonexistent prior to the gauging process. Public opinion is defined as the average judgment or consensus of the individuals of a "public" regarding a given issue, institution, or person.

---

* This critical review was undertaken at the request of a number of social scientists who believed that a critique of the opinion-attitude methodologies should be made, preferably by someone without prior or vested interests. The writer, having reluctantly agreed to attempt the task, hopes that he has singled out for discussion the more significant of the many issues in this expanding area of research. Acknowledgement is due to Paul R. Farnsworth for suggestions from time to time and for a critical reading of the manuscript. Olga W. McNemar has assisted in the rather heavy library work involved in scrutinizing about 800 references, and she has also served as the writer's critic during the preparation of this report. I am indebted to many individuals in governmental agencies with whom I had contact during the two years beginning with the autumn of 1941 and especially to those consulted during the winter of 1944. I am grateful to the Social Science Research Council for a grant which facilitated this appraisal.

Some writers use the terms attitude and opinion as having very similar if not the same meaning; both are interpreted as having to do with a predisposition to action. Thus conceived, one can have an unexpressed opinion; if it is expressed, it may be taken as one's opinion or as evidence regarding one's attitude. Since the terms are so frequently used synonymously, and since what one investigator calls an opinion study might be called an attitude study by another, and vice versa, we shall not attempt herein to adhere to any rules about the noninterchangeability of these words. A useful, albeit somewhat artificial, distinction can be made on the basis of techniques: *The typical attitude study involves a scale or battery of questions for ascertaining attitudes whereas the typical opinion, particularly public opinion, study leans heavily on a single question for a given issue.* Although the two have much in common, there is sufficient difference to permit separate discussion of certain aspects. Section II of this critique will be devoted to "measuring" attitudes by the use of scales, and the single question method will receive attention in the following section.

In brief, the several techniques used for getting at opinions or attitudes boil down to the simple matter of asking people questions about an issue in order to elicit a response which is interpreted as the respondent's opinion about or attitude toward the given issue. The introduction of high sounding definitions does not alter the fundamental fact that practically all attempts at determining opinion and attitude are at the verbal level—the correlation between verbal and overt nonverbal behavior, the latter dependent in part upon the opinion or attitudinal set, is an unknown and is usually left as an unknown by investigators in this field. This is one aspect of the question of validity to which we shall return presently.

The form of the questions asked by interviewers or in schedules varies greatly. Some call for a "yes" or "no" answer; some may require a rating such as "very much," "much," etc; sometimes a list is provided and the respondent is asked to check or rank the items; occasionally a preference between two possible alternatives is called for; quite frequently the question is of the open-end or nondirective type—the respondent is free to phrase his own response; at times the respondent is presented with a statement or series of statements and asked to indicate whether he agrees or disagrees therewith. Still other variations have been used.

Regardless of the question form and how the answers are obtained, the responses are used to categorize the individual as belonging to one of two or more classes. These classes may represent qualitative or near-quantitative differences. The category to which an individual's opinion is assigned may correspond exactly to his response or responses; it may depend upon a rating by the interviewer or by the analyst who studies and codes the verbatim interview; or it may depend upon either an arbitrary or rational scoring of the responses to several items. This last procedure is an attempt to assign to an individual some position or rank on a continuum or near-continuum so as to characterize his belief in, agreement with, or feeling about the given issue. What questions or items should be used for this purpose and how they should be weighted is the problem of "scaling" to be treated in Section II.

The studies, which are often characterized as surveys, are always done on a sampling basis. The aim is to secure a sample which, within the limits of random or chance errors, can be thought of as a representative cross section of some defined population. Thus the inferences made or generalizations deduced hold only for the population sampled, and any inadequacies in the sampling procedure should, perforce, impose qualifications on the conclusions. Most of the so-called attitude studies have been made on college student groups, or others which happen to be available. Usually the investigator fails to specify how he drew his sample, and from what population. Sampling issues will be briefly discussed in a later section.

The statistical analysis of the data of opinion and attitude research may be simple or relatively complex: simple if the aim is merely to report an over-all percentage; more complex if the aim is to make group comparisons, to correlate opinions with other variables, or to study the interrelatedness of opinions. The problems of analysis are not much different from those in other fields of social science. Pitfalls and hazards exist and, as will be seen in later pages, some errors and inadequacies in statistical treatment occur rather frequently.

The foregoing sketch does not include all the methodological issues which are to be treated in subsequent pages. There are certain questions concerning the securing of the data—by direct interview, by mailed questionnaire, by group administration of schedules, etc. The study of opinion-attitude changes is of sufficient importance to deserve somewhat detailed exposition. The same is true of the determination of correlates and intercorrelates of attitudes. Some other persistent methodological problems are illustrated in Section VIII.

Fields of Application. The reader who is not already familiar with the rather extensive use of attitude-opinion techniques may find a few paragraphs thereon of interest. It is not enough to say that these methods have been used by journalists, sociologists, political scientists, psychologists, educators, consumer research specialists, *et al.* What we wish to convey here is some notion of the diversity of more specific fields of application. It will become quite obvious that the boundaries of disciplines are immaterial. It is difficult to say to which group belongs the major credit for developing the techniques, but it is reasonably certain that the psychologists and sociologists are mainly responsible for the current status of attitude research (by scales), and that the journalists have contributed much to opinion research (by single questions).

All readers are familiar with the Gallup and *Fortune* polls and the types of issues—political, economic, international, etc.—dealt with primarily as a service to journalism. The main aim of these polls is the reporting of opinion on various timely topics, and the determination of shifts of opinion regarding certain issues. Usually over-all percentages or percentages for a few breakdowns are placed on record. The results of these surveys generally appear only in newspapers and magazines, seldom in scientific and scholarly journals except for the periodic appearance in the *Public Opinion Quarterly* of current opinion surveys carried out by *Fortune*, the American Institute of Public Opinion (Gallup), and the National Opinion Research Center (Denver).

In the professional journals hundreds of articles on attitudes (mainly) and opinions have appeared during the last fifteen years, with additional hundreds prior to this period. In general, these more remote papers are of less importance than those following the impetus given research in this field by L. L. Thurstone about 18 years ago. By Thurstone or under his direction scales were constructed for "measuring" attitudes toward the church, treatment of criminals, capital punishment, God, communism, patriotism, the Negro, birth control, the Constitution, the Bible, Sunday observance, war, evolution, censorship, the Chinese, the Germans, etc. A few years later a series of scales appeared under the authorship of H. H. Remmers and his students. These were so-called generalized scales designed to measure attitudes toward any social institution, any school subject, any teacher, any dramatic production, any disciplinary procedure, any practice, any racial or national group, any proposed social action, any selection of poetry, any vocation or occupation, any homemaking activity, in fact nearly *any* anything. In order to pave the way for grinding out more scales of this type, one graduate student (**100**, pp. 117–126) was set, or found, the task of building up lists of the multitudinous miscellany about which people can and maybe do have attitudes.

Attitudes toward rural and urban life, the social rights of Negroes, many economic policies, the T.V.A., relief, and the public press have been studied. Attitudes regarding law, juvenile delinquency, behavior problems, and various aspects of religion have been investigated. The attitudes of children toward parents and of parents toward children have also been tackled. Education, nursery schools, the high school, and intercollegiate athletics have been the object of attitude research. Opinions about radio advertising, the movies, retail stores, women executives, feminism, the Dies Committee, political parties, marriage and sex have been collected and analyzed. Many publications have been concerned with that complex of attitudes and opinions having to do with religious, political, social, and economic liberalism-conservatism.

The various techniques have been utilized for the study of job satisfaction, employee attitudes, several aspects of the morale of civilians and soldiers, optimism-pessimism, and satisfaction with life. Attitudes toward the future and toward death have not escaped notice.

It can be said that some of the questionnaires used in these surveys contain everything but the proverbial kitchen sink, and once such a questionnaire has been filled in by a sizable group its author has the "basic" data at hand for a half dozen articles. If he is fortunate enough to have punched card equipment, it becomes the misfortune of his professional contemporaries to find the literature being filled with results of cross tabulations which are so lacking in rationale as to be nonsensical. The "hypothesis" step in scientific reasoning and research seems to be all too frequently ignored by the users of these techniques. If by 1931 a whole book (132) could be devoted to the attitude field, one might expect that from then on enough facts about attitudes would have been known to have permitted systematization sufficient for fruitful theories and hypotheses.

Some Fundamentals. It may be advantageous next to discuss briefly a few of the fundamental requirements which all techniques for "measuring" attitudes and opinions must meet. Certain of these are so basic that unless reasonably well attained it cannot be expected that the study of attitudes and opinions will ever approach scientific status. As will be seen presently, the attitude scale or battery of questions and the opinion or single question methods differ somewhat in the extent to which an attempt is made to meet these basic requirements. If one is to judge from current practices, it can be inferred that some attitude-opinion researchers will brush these requisites aside as being old-fashioned.

The presumption, of course, is that attitudes and opinions are being measured or gauged by use of these methods. The extent to which the terms "measured" and "gauged" are applicable, or just what these terms can connote, needs to be scrutinized. If one is attempting to measure or gauge something, the accuracy or error of measurement must be known. This problem of reliability is common to the physical, the biological, and the social sciences. A very large number of the variables of science are measured indirectly, and such indirect measurement raises the question of validity—the extent to which the instrument is measuring the variable it is designed to measure rather than reflecting some other variable or variables. It is nearly always desirable that a given instrument should measure just one dimension, i.e., involve one continuum.

Measurement. What can this term mean in opinion-attitude research? A number of critics of psychological measurement in general, and attitude measurement in particular, have rightly claimed that such is not measurement in the true sense because nothing is known concerning the equality of the units used in the scales. At first thought, this criticism might seem quite disturbing, but actually it only means that certain limitations of these so-called scales must be kept in mind. If we have A scoring 4, B scoring 6, and C scoring 8, it simply cannot be said with any certainty that A and B differ as much as C and B or that C possesses twice as much of the attitude as A. What can be said is that B's value differs from that of C in the same direction that A's value differs from B's, a statement which assumes that a single continuum is involved. In order to be sure that C is more favorable in his attitude than B who in turn is more favorable than A, it would be necessary to know the possible magnitude of the "error of measurement."

It should be understood from the outset that no claim is made herein to the effect that attitude measurement permits more than a rank ordering of individuals. The single question or opinion gauging technique does not accomplish even this much, but it does permit a rank ordering of groups of individuals.

Reliability. In discussing the concept of reliability it is necessary to distinguish between two current usages of the term. Its more general meaning seems to be prevalent among the opinion pollers, whereas a restricted meaning seems to have been accepted by the attitude testers. In its more general sense it means the dependability of a result—thus encompassing not only the restricted, or accuracy, idea but also the notion of validity and sampling adequacy. By reliability we shall mean the accuracy with which an individual's attitude is measured, or the degree of error involved in assigning an individual to a class or in establishing his rank-order position. This is the commonly accepted usage among psychologists. Reliability in the sense of sampling dependability will be subsumed under sampling stability of statistical measures. The term validity will also be used in a restricted sense: Does the test, scale, or question tap the variable it is supposed to measure? Thus we shall *not* speak of the reliability of a percentage or of the validity of a study.

All measurement is befuddled with error. About this the scientist can and does do something; he ascertains the possible extent of the error, determines whether it is constant (biasing) or variable, or both, and ever strives to improve his instruments and techniques. Physical scientists usually determine the magnitude of measurement errors by making several measures on the same object, then expressing the error as the average error (in terms of original units) or as the percentage error (the average error relative to the object's magnitude).

In the psychological measurement of a trait or characteristic of a person, it is impractical to make more than one or two determinations for a single individual, hence the amount of error cannot be found in the same manner as in the physics laboratory. This apparent predicament is overcome by taking two measures on a large number of individuals, say 100 or more. Although the average discrepancy could be taken as indicating the error, it has been found more convenient to compute the product moment correlation coefficient between the two sets of measures, which sets are assumed to be comparable measures of the same characteristic. It is well known that from this reliability coefficient and the standard deviation for the trait as measured, one can state the "standard error of measurement," which is analogous to the "average error" in that it is in the units of the test or scale. Relative error in the physicist's sense is inapplicable here because of the lack of a real zero point, but a relative measure is available in the reliability coefficient itself which is inversely proportional to the error variance *relative* to the trait variance.

It is, of course, well known that the nearer the reliability coefficient is to unity, the more accurate the measurement. Since perfect reliability is impossible, and particularly difficult even to approach in psychological measurement, the question is frequently raised as to how high the reliability must be for a given purpose. To this there is no answer which will be satisfying to all. The writer is of the opinion that, if a study involving measurement is worth doing, every effort should be made to attain accurate measurements rather than being satisfied with just any reliability. The writer is also of the opinion that the 18-year-old dictum of one of our better known statisticians to the effect that a reliability coefficient of .50 is sufficient for group comparisons has been a disservice to the cause of adequate measurement. Far too many investigators have hidden their heads in the quicksand of that dictum. One of the characteristics of mediocre research is mediocre reliability of the instruments used.

The specific reliabilities attained by attitude scale techniques will be given in the next section and what little is known of the reliability of the single question method will be presented in the third section. Suffice it to say here regarding the latter that the accuracy with which one is classified as, for example, favorable or unfavorable to a given proposition can be ascertained even though no expression similar to error of measurement is possible. A type of correlation can be calculated providing two responses are obtainable from each of several individuals. These two responses can be to two different questions designed to get at the same opinion; or a second response to one question can be secured at a later date. This is not the place to discuss which of these two procedures is the better, or to elaborate on the difficulties involved in either scheme. It may be noted, however, that some have objected to the second scheme because it is feared that bona fide changes may have taken place; thus an individual would not tend to give the same answer twice. This poses a dilemma: If an opinion or attitude is such a momentary thing that individual reliability or consistency is lacking, how can one expect the opinion in question to correlate with stable sociological or psychological characteristics of the individual? The research problems associated with unstable as opposed to relatively stable opinions and attitudes must be quite different, yet we fail to find this distinction in the literature.

The factors which lead to variable errors are well known, and need not be catalogued here. The nature of possible constant errors as an aspect of individual reliability differs somewhat for the attitude scale and the opinion question. If a neutral position on some issue is of importance per se or as a point of reference, then the selection and weighting of the questions for a scale and the wording of the single opinion question have an important bearing on biasing the score or the response. But in case of an attitude scale having no absolute neutral point, it is difficult to see what bias could mean. A constant error of four points would only result in shifting the average, which average has no absolute value anyway. For the single question with or without a provision for a neutral response, bias can of course be introduced, perhaps never avoided, but such bias would not act as a constant error since not all individuals are affected. The biasing error definitely affects some, and therefore the reported over-all percentage for a group will be biased. Variable errors, however, would not lead to a predictable error in a percentage, but would add to the variable error already inherent in the percentage as a result of sampling.

Validity. The problem is essentially that of supplying evidence that the device used measures or classifies the attitude or opinion it was designed to measure. This is an old problem which has engendered much debate, and from which many avenues of escape have been attempted. Does an expressed opinion actually represent a person's real attitude? Can verbal or symbolic behavior be depended upon as indicating an individual's action tendency? What is the degree of relationship between overt nonverbal and verbal behavior? The problem is sharpened by reference to the work of La Piere (**70**) who found that practice and verbal behavior on a race issue were markedly inconsistent.

Validity can be ascertained by a variety of methods, which will be listed here without much comment. Sometimes verbal behavior can be checked directly against nonverbal overt behavior, and a correlation thereby established. Evidence for validity can be obtained by comparing those of known attitude as inferred from their activity or voluntary membership in organizations. Presumptive evidence can be secured by learning whether the scale or question differentiates between groups which on a priori grounds should differ in their opinion about the given issue. Sometimes it is possible to utilize the ratings of close acquaintances as a criterion for checking validity. All too frequently a new scale is checked against an older scale, the assumption being that a high correlation between the two indicates validity. This practice has been overworked by the mental, the attitude, and the personality testers who are seldom too careful about first being sure that the original scale's validity has been established. Another scheme which might aid in ascertaining validity would be to follow the scale administration or the single question by intensive and extensive interviews in order to see whether the first expressed position holds up under cross examination.

It should be noted that any of these schemes for getting at validity can be followed out on a miniature cross section of 100 or more cases. All of these schemes involve information external to the scale itself. Opposed to the use of external criteria is the a priori method of validating. A scale for attitude toward an issue is said to be valid because the items in the scale make it so, that is, the items have been judged valid. This does not overcome the fact that the verbally administered scale calls for verbal responses which may or may not reflect the individual's attitudinal set. The same is true when an opinion is elicited as a response to a single question. As an illustration, let us look at some results reported by Corey (**27**) on attitude toward cheating. A scale to measure this attitude was constructed by one of the standard procedures. The reliability coefficient of .91 is adequate. When scale scores were correlated with actual cheating behavior, the relationship was an insignificant correlation of .02. Evidently the scale did not tap the action tendency for which it was designed. One cannot, of course, always be sure that attitudes can be correctly inferred from overt behavior—it is possible for such to be an act, or conforming behavior. A study by Schanck (**114**) has led to an attempt to distinguish between privately held and publicly expressed opinions. For example, at church a man will publicly take a stand against card playing although he privately admits that he favors card games.

Some investigators have sidestepped the problem of validity by denying that anything exists beyond the verbal expressions, hence there is no problem of validity. Others have adopted the idea that scales or questions test whatever they test, so why worry.

A word should be said about item validation via internal consistency. It is the writer's opinion that internal consistency has to do with reliability and that validity cannot be inferred therefrom.

There are several obvious factors which affect validity and/or reliability. It is difficult to see how a reliable and valid response can be secured unless the respondent understands the given issue. Likewise, he must be able to understand the questions asked him. Ambiguous wording must be avoided. Other things being equal, the less the personal relevance of an issue, the lower the reliability and validity. Stereotypes and emotionally charged words or phrases, though possibly leading to greater consistency, are not conducive to high validity and reliability. Some of the disturbing factors in the wording of statements and questions for attitude scales can be minimized by following the rules set forth by Wang (**125**). The problems involved in the wording of questions for opinion polling will receive attention later, as will also the influence of interviewing as contributing to possible erroneous responses.

In connection with the subject of validity, it might be noted that the absence of any objective evidence for the validity of a new scale does not dampen the tendency to give it a name. Once named, its author and other users all too frequently find it convenient to forget that its validity is unknown. A still worse practice is the tendency of some to rename a scale which was originally constructed as a measure of a particular attitude, and thereafter treat it as a measure of some other attitude. For example, scales constructed (validated?) by the Thurstone technique to measure favorableness of attitude toward God, the church, and war, have been interpreted as measuring liberalism-conservatism (**63**). Perhaps this is true but where is the evidence, and what does liberalism toward God mean?

Single continuum. Measurement implies that one characteristic at a time is being quantified. The scores on an attitude scale are most meaningful when it is known that only one continuum is involved. Only then can it be claimed that two individuals with the same score or rank order are quantitatively and, within limits, qualitatively similar in their attitude toward the given issue. As an example, suppose a test of liberalism consists of two general sorts of items, one concerned with economic, the other with religious issues. Two individuals could thus arrive at the same numerical score by two quite different routes. Now it may be true that economic and religious liberalism are correlated but, unless highly correlated, the meaning of scores based on such a composite is questionable.

This leads to the problem of general versus specific attitudes, a controversy which can best be settled by resort to further investigation. At this place we are only trying to point out that one fundamental requirement for an attitude scale is that it place individuals in rank order along a single continuum. The intelligence testers have gone through a cycle: first the measurement of simple functions, then of more complex and general functions by a hodge-podge of items, and recently, partly as a result of factor analysis, the stress is on the development of tests, each of which involves a single continuum or a so-called unitary ability. Attitude testers have gone through the first two phases of this cycle, but some are still willing to tolerate a morale attitude score made up of four uncorrelated components! This haphazard adding of dissimilar parts on the assumption that a meaningful whole will result is nonsense.

Upon first thought it might seem that this single continuum requirement need not be considered when individuals are to be classified by the single question method. It is true that no continuum is apparent, unless one is inferred as underlying the expressed opinion, but a similar requirement does exist in that it is hoped that those who are classified as favoring an issue are relatively homogeneous. To be classed together should mean similarity, and this can be attained, at least approached, by the use of a question which is unambiguous, which imposes a similar reference point on all respondents, which is so easily understood that personal interpretation as to the meaning of the question is ruled out, and which leads to an unqualified answer.

The single dimension requisite is not only pertinent when a score or response is being interpreted; it has an important bearing on any research which is designed to analyze individual differences in attitude, i.e., to determine the correlates of attitudinal differences. Such an attempt at "explaining" variation is handicapped if the variability under analysis is rendered more complex by the presence of many continua in the measuring instrument. One does not, of course, get rid of all the complexities of social psychology by isolating unitary characteristics and developing "pure" tests thereof. This is only a step in the direction of avoiding the one complexity introduced by failure to construct scales for measuring one dimension at a time.

Another aspect of the single continuum problem is the meaning of the two extremes on a scale. If one end indicates a favorable attitude, does the other end represent unfavorableness or something else? For instance, is Fascism the opposite of communism? Such would seem to be the case in the minds of some. Are we dealing here with one or two dimensions? To interpret the extremes as being real opposites or to interpret the extremes as on different continua is precarious in the absence of knowledge concerning the real meaning of the extremes.

These basic considerations regarding units of measurement, reliability, validity, and dimensionality will again come to our attention at appropriate places during the following detailed discussion of attitude and opinion measuring techniques.

II. Attitudes by Scaling Techniques

Pursuant to our rather arbitrary distinction between attitude measurement and opinion gauging, we shall now consider the methodological problems involved in the construction of scales which permit rank ordering of individuals into a number of graduated classifications. For convenience, scaling methods are being separated from the single question or polling method.

The aim of scaling is the development or construction of a "measuring" device which will distribute individuals along a continuum running from a highly unfavorable through neutral to a highly favorable attitude. The number of steps or gradations, sometimes called discriminability, which is required will depend upon the fineness with which one wishes to differentiate between individuals. Other things being equal one can assume that the finer the possible gradations, the more reliable the instrument, but if one can secure satisfactory reliability with, say, 50 possible scores or positions, there would seem to be no reason for seeking greater discriminability. In so far as the likely statistical treatment of results is concerned, it would seem that from 12 or 14 to 20 possible (and actually found) gradations should be sufficient. This follows from the well-known fact that statistical terms such as means, standard deviations, and product moment correlation coefficients are not seriously in error when computed from coded values providing that at least 12 or 14 code values are used, and the gain from having more than 20 is negligible. Additional possibilities should be considered. In the first place, the fact that, say, 16 steps are available does not guarantee that the scores will fall over a range of 16 values and, secondly, even if the possible range is utilized in one group, the instrument may not permit fine enough classification when groups of greater homogeneity for the given attitude are being studied.

The assumption that individual differences in a given attitude can be thought of as falling on a continuum has brought forth vigorous objections. It is said that this may or may not be the case. Suppose that it has been demonstrated that a certain attitude does not lie on a single continuum, it does not follow that measurement is impossible. Instead of trying to place individuals along just one continuum, it becomes necessary to develop scales for as many continua as have been demonstrated to exist. How many is an empirical problem, as is also the related question of the extent to which attitudes are general as opposed to specific. The burden of proof as to whether an attitude as measured does follow a single dimension falls upon the one who constructs the measuring scale, or develops the scale construction technique. The requirement that a scale shall reflect a unitary attitude does not rule out the possibility of attempting to measure general attitudes or complex attitudes or specific attitudes. The complex ones would need to be broken down into component parts and a scale provided for each part. It may be that some so-called attitudes will not be amenable to scaling as unitary characteristics.

The early efforts at attitude or opinion measuring usually involved a questionnaire or battery of questions which were selected on an a priori basis. Numerical values were assigned arbitrarily to the item or question responses and these values were summed to secure scores which were then interpreted as indicating the attitude of the respondents. There was nothing about the procedure to guarantee that any one item tapped the same attitude as the other items. Little thought was given to the number of dimensions—that more than one was frequently involved seems obvious. Reliability was occasionally determined, but validity was seldom mentioned.

An attempt to improve on this haphazard work was reported in 1925 by F. H. Allport and Hartman (1) who reasoned that the questions or statements to which responses were to be secured should be so chosen as to represent varying degrees of favorableness or unfavorableness. Their report came to the attention of L. L. Thurstone, who saw that the assignment of statements to a favorable-unfavorable continuum was a problem amenable to solution by various psychophysical methods, the most feasible of which was the method of equal appearing intervals.

A detailed discussion of the Thurstone attitude scaling technique need not be given here; any reader who is unfamiliar with the method can turn to a number of sources, particularly a 1928 paper by Thurstone (**122**) and the monograph by Thurstone and Chave (**123**). Suffice it to say here that the scheme is one for arranging items on an 11-point scale according to the degree of favorableness or unfavorableness as determined by having a large number of judges sort quite a number of statements into 11 piles. The median judged location for an item is its assigned scale value. Objective criteria were developed for eliminating items which are ambiguous and which are irrelevant in the sense of not belonging to the continuum from the viewpoint of the judges as sorters. The process, of course, begins with the assembling of a much larger number of statements than needed in order to be able to select, from those not eliminated on the basis of ambiguity and irrelevancy, a desired number, say 20, with scale values nearly evenly spaced along the assumed continuum. With sufficient statements, two comparable forms can readily be constructed.

In taking the test, the respondent is instructed to check those statements with which he agrees, and his score is the median of the scale values of the checked statements. Such scoring permits a very large number of score positions. The reliability coefficients computed by the form versus form method for scales so constructed, for a variety of attitudes, are usually between .70 and .90, with typical values in the low .80's.

That the scales may possess validity has been demonstrated by using groups of known (?) attitudes, but just how high or low the validities are has not been adequately determined.

There have been numerous and varied criticisms of the Thurstone method, many of which would have been forestalled by a careful reading of Thurstone's publications. All attempts to disprove his working assumption that the determined scale values will be independent of the attitude of the sorters toward the given issue have been futile. The claim that higher reliability would result if the mean, instead of median, scale value of the checked items were used in scoring seems unsubstantiated by research on this problem by Lorge (**81**).

Some have been critical of the units along the base line or continuum, claiming that these units are not equal. The present writer has been unable to find that Thurstone ever claimed equal units for his scales—he has spoken of equal *appearing* intervals. That any scheme utilizing statistical measures based on a sample can ever lead to equal units would seem impossible because of the presence of sampling errors. Thurstone has claimed that scales constructed by his method for different attitudes would permit the direct comparing of scores as between two attitudes To this there can be no objection providing a strict comparability of scores is not assumed. The writer has not been able to see that the defined neutral points for two different attitudes are really the same. **If** the average score for the attitude of a group toward capital punishment is 6.5 and toward law is also 6.5, can it be said that the group averages as far from a neutral position on one issue as on the other? The type of comparability of scale values which has been rather generally accepted in other fields of psychological testing is that percentile ranks, or standard scores for normalized distributions, denote similar positions. If such is the case, then scales by the Thurstone technique definitely do not yield comparable units; if the Thurstone method does lead to similar metrics when applied to different attitudes, then the intelligence testers err in assuming the comparability of percentile ranks. Attitude and aptitude variables are, of course, quite dissimilar but when "measured" one notes the common element: individuals are assigned scores which are supposedly indicative of the degree of ability in one case and the degree of favorableness in the other.

It is the writer's opinion that the comparability of scales constructed by the method of equal appearing intervals has not been sufficiently well demonstrated to permit the direct comparison of means based on different scales. Remmers and followers (**99**, **100**, **101**) have been the most persistent in this practice.

A recent study by Riker (**104**) may have some implications for questions regarding the neutral point, and the comparability of values. He used six scales which had been constructed by Thurstone, or under his direction, and 218 Princeton undergraduates who were asked to check their degree of favorableness-unfavorableness on an 11-point graphic self-rating scale and also their intensity of feeling on an 11-point self-rating scale. Thus Riker had distributions for Thurstone scaled scores, and for two types of self-ratings for attitude toward the Negro, the Germans, the treatment of criminals, capital punishment, evolution, and communism. He reported that the largest difference between means for Thurstone versus rating scale values is .916 times its standard error, hence the differences could easily be chance. He concluded from this that the three methods yield "similar metrics." If this were so, we would have some confidence that the neutral points for the three are comparable, but it turns out that Riker made some kind of a consistent error in determining standard errors. Our own calculations, with which corrections recently published by Riker (**105**) agree, indicate that the critical ratios range from .7 to 19.4. Enough (15 of 18) of these critical ratios are sufficiently high to lead us to conclude beyond any doubt that these three schemes do not yield "similar metrics," a fact which also should have been apparent to Riker had he regarded the failure of the three methods to yield similar standard deviations.

It follows, therefore, that a group with mean scale score at the defined neutral point will not necessarily rate themselves as neutral in attitude. It may be that the neutral point on a scale constructed by the Thurstone technique does correspond to "real" neutrality on a scale of absolutes, but this has not been definitely proven. The writer would grant that the presumptive evidence favoring the assumption is far greater for the Thurstone type of scale than for the less rational scales for which the neutral points are arbitrarily designated as the mid-point of the possible scoring range.

Kirkpatrick and Stone (**68**) have criticized the Thurstone method on five counts, which actually boil down to three: inequality of units, questionable assumption of attitude continua, and ambiguity due to different motives or reasons back of responses to different items. The matter of continua has already been discussed herein. As to inequality of units, the criticism is that the interchangeability of units on the base line has not been demonstrated. The same criticism has been made more forcefully by others, particularly by Johnson (**62**) and Merton (**86**). With the criticism we agree (see Section I), but we question whether any progress is made in Kirkpatrick and Stone's solution to the problem. They proposed that items which had been retained on the basis of the judges' agreement (1) on whether the item was favorable or unfavorable, (2) on the point of reference as social, personal, cosmological, or epistomological, and (3) as to "vigor of wording," should be simply scored as plus one and minus one for pro and con respectively. They reasoned that "At least in the method of scoring used there is a counting process and a very simple manipulation of multiples of units which are relatively interchangeable," and hence concluded that a quantitative variable had been derived. They have not shown that any two statements lead to responses which are even relatively equal so as to make the score of "plus one" on one item equivalent to a "plus one" on another. Counting apples is not measurement even though very simple manipulation is possible. It would seem that the real meaning of interchangeability of measurement units has not been understood by these authors. Elsewhere Kirkpatrick (**67**) says that inter-person interchangeability is lacking. Is that what the physical scientists mean by interchangeability of units?

It is also of interest to note how Kirkpatrick and Stone handled the problem of ambiguity of reference point. As indicated above, items were not retained unless there was agreement as to the basis of argument or point of reference. As we understand their report, they did retain items which involved at least four different reference points, then proceeded as though they had little faith in their own valid criticisms regarding the use of questions which may be responded to from different viewpoints.

It is the conviction of Merton (**86**) that since Thurstone's method does not lead to units which are additive and interchangeable on a linear continuum, the scales should be abandoned in favor of item analysis, which he considers "one way in which opinionnaire results can be legitimately employed without recourse to dubiously applicable mathematical operations." He reasons that for a true scale, persons endorsing an item graded as very unfavorable should endorse all statements which are slightly less unfavorable. Ideally, yes, but Merton who says that the advocates of scaling are too far from the "facts of sociology" ignores the just as well established psychological fact that the responses of individuals to nearly all types of questions are subject to response errors. These errors not only complicate the problems of true scaling but they also disrupt analyses based on single items; the latter fact is not mentioned by Merton. As will be seen in the next section, there are many factors which make the percentages based on single questions or statements undependable.

Since the ever present response errors are such that an "ideal" scale can never be realized, one wonders how much of a digression from the ideal would be tolerated by Merton. Perhaps he is under the false impression that measurement ever reaches an ideal of perfection. We hasten to add that there is ample evidence to support Merton's position that one of the chief defects of the Thurstone scales is the scattering of the endorsed statements over a wide range. In Dudycha's 1943 study (**32**) of the Peterson attitude toward war scale, it was found that those scoring most unfavorable toward war too frequently endorsed statements with scale values at the opposite end of the continuum. Dudycha concluded that the presence of several different continua would explain this scatter of endorsements. He might have added that the spread was probably associated with the none too satisfactory reliability of scales constructed by the equal appearing intervals method. That more than one continuum may be involved in scales constructed by this method is suggested in the report (abstract) by Ferguson (**38**) who found by factor analysis that more than one component is necessary to explain the item intercorrelations on the war scale. Detail for evaluating this study is not available.

A 1942 paper by Ericksen (**33**) has been cited by Dudycha as indicating that something is wrong with the Peterson war scale because it did not reflect group differences and changes which seemed logical, and which item analysis seemed to bear out. Ericksen's failure to find statistically significant differences between groups was due to erroneous computation of the standard errors of the differences between means. Instead of using the standard errors of the separate means, he used the standard deviations of the score distributions; hence all of his nine reported critical ratios tend to be quite low, .32 or less. With the use of the proper formula, some of the logically expected differences are statistically significant. Since a part of Ericksen's criticism is based on changes in responses to three supplementary questions as contrasted with the supposedly insignificant change in scale scores, occurring between May 1940 and February 1941, it is of interest to look at the three questions which indicate "a marked increase in the militaristic feeling between the two testing periods." The questions were: "Do you think the U. S. will enter the present war?" "Do you think the U. S. should aid the allies in every way except by sending troops?" "Should the U. S. fight to protect the Western Hemisphere from a German invasion?" One might grant that a die-hard pacifist would respond negatively to these questions, but does an affirmative response really indicate militarism? Considering the events between May 1940 and February 1941, does a change to more yeses to these questions indicate greater militarism or just good old fashioned realism regarding defense? It may be that Ericksen's general conclusion that the Thurstone type of attitude scale "is too general and abstract in its reference to be practical or valid as a measure of contemporary specific war attitudes" is justified, but not on the basis of his data.

As in all attempts to elicit verbal responses, there are certain rules to be kept in mind when statements are being prepared for possible inclusion in an attitude scale. The interested reader will find the paper of Wang (**125**) quite useful. The tedious sorting work involved in the method of equal appearing intervals can be shortened by a scheme set forth by Seashore and Hevner (**115**) or by a graphic rating method proposed by Ballin and Farnsworth (**2**). That scale values determined by the Thurstone technique may shift has been shown by Farnsworth (**37**) who found significant changes for items originally scaled in 1930 and rescaled in 1940. Even though the rank-order positions of the two sets of values correlated to the extent of .97, the displacement was such as to produce a difference in means for 256 cases scored on the basis of the old and new values. The difference was .69, which yielded a phenomenal critical ratio of 34. Evidently there may be some danger in assuming the same meaning for scores from year to year.

The Thurstone scaling method as used by him and his students has always involved scaling for some particular attitude. Remmers and Silance (**102**) proposed in 1934 that the method could be adapted for the purpose of constructing "generalized attitude scales." Beyle (**4**) had made a similar proposal in 1932. This assumes that a series of statements can be scaled as applicable to a class of objects or phenomena, such as races, or school subjects, or proposed social changes—the "any" type of attitude listed in the first section. Thus one sorting and subsequent scaling provides a set of statements which serve for measuring attitude toward, for example, any institution such as war, marriage, communism, Sunday observance, and divorce. The advantage of the generalized scale is supposed to reside in the fact that many attitudes can be measured without the great expenditure of time otherwise required for sorting and scale construction.

To make the scale applicable to any one of the phenomena in a class, the respondent is simply told that he is to keep in mind the particular attitude object as he checks the statements. That this procedure can become quite ridiculous is illustrated by the following items: "[war or marriage] does not consider individual differences"; "[marriage] gives too little service"; "[marriage] satisfies only the most stupid with its services"; "[marriage] offers opportunity for individual initiative"; "[war] is too conservative"; "[war] is too changeable in its policies"; "[marriage] appeals to man's lowest nature" (**99**, pp. 18–36).

In order to establish the validity of the generalized attitude scaling technique, one of Remmers' students (**99**, pp. 98–109) constructed a scale specifically for attitude toward teaching and then correlated the scores therefrom with scores based on the generalized attitude toward any occupation checked for teaching. The correlation corrected for attenuation was only .58, but this is said to "show a satisfactory validity" for the generalized scaling scheme even though the validity of the specific scale is unknown. Remmers and his students (**99**, **100**, **101**) have practically placed attitude scaling on a mass production basis, and have proceeded to use the scales as tools in research on a variety of topics. Individual citation of these studies would add more than 40 titles to a bibliography on attitude research. The fact that the dozens of reliability coefficients reported tend to have a median value of about .70 with values as frequently below .50 as above .80, and ranging down to .07, would indicate that a great deal of uncritical energy has been expended. These reliabilities are lower than those usually found for the Thurstone scales even though from 50% to 100% more items are included in the generalized scales. Evidently the latter scales contain dead timber.

In 1932 Likert (**77**) published a paper in which he presented a method for developing scales for attitude toward internationalism, the Negro, and imperialism. Although this is not a rational approach in the sense that Thurstone used the term rational, the end results are sufficiently satisfactory to justify a sketch of the technique used by Likert. Items or statements which call for checking one of five responses (strongly approve, approve, undecided, disapprove, strongly disapprove; agree-disagree may be substituted for approve-disapprove) were prepared on the basis of a priori judgment. At first, sigma scoring weights for the possible responses were determined by converting the proportion of times each response was checked into the corresponding sigma value on the base line of the unit normal curve. That this laborious conversion job and the resulting cumbersome weights were unnecessary became quite evident when the simple, though arbitrary, scoring weights of 1, 2, 3, 4, 5, for the respective responses were tried out. Scores by this simpler method yielded correlations with scores by sigma scoring which were in excess of .99 and, as one would anticipate from this fact, there was no noticeable difference in reliabilities.

The final selection or elimination of items did not depend upon subjective judgment, but upon the criterion of internal consistency, that is, each item had to show appreciable correlation with the total score on all the items of a scale. The three final scales contained 24, 15, and 12 statements, and the odd-even reliability coefficients ranged from .79 to

.92 while test-retest (30 day interval) reliabilities ranged from .79 to .91. Considering the shortness of the scales relative to the usual 20 to 22 statements in the typical Thurstone type scale, these reliabilities are fairly satisfactory. Eighteen of the 22 statements of the Thurstone-Droba war scale could readily be responded to in terms of the five agree-disagree responses mentioned above. Scoring on the 1, 2, 3, 4, 5 basis increased the reliability to .88 as compared to .78 secured by scoring all 22 items by the Thurstone method.

In further study of the possible advantage of this scoring scheme, Likert, Roslow, and Murphy (**78**) administered 10 Thurstone scales to several groups. The subjects were instructed to check with a + the statements with which they agreed and to circle the + if they strongly agreed (this much permitted scoring by the usual median check method of Thurstone); the statements with which they disagreed were marked by a −, which was circled if they strongly disagreed; if undecided, they used the question mark. All blanks were then scored by the two methods. The form versus form reliability coefficients were higher for each of the 10 attitudes for all the groups when the scoring was by the Likert method. Actually the reliabilities were appreciably higher—in the .90's as compared with the .80's. Furthermore, the correlations obtained between scores by the two methods were as high or nearly as high as the respective reliabilities would permit. This study would seem to be conclusive evidence in favor of the Likert simple scoring method. Nowhere does it involve scale values (dubious to some) as determined by the method of equal-appearing intervals. Does such a scheme constitute measurement? Yes, providing it can be shown that a single continuum is present and providing it is understood that by measuring we mean only rank ordering of individuals.

The foregoing discussion, of course, does not provide grounds for concluding that the Likert technique is superior to Thurstone's scaling as a method for *selecting* items. That the reliabilities for scales built by the Likert technique do not greatly, if at all, exceed those attained by equal appearing interval scales is evident from the reliabilities of .79 to .92 reported by Likert for his own three scales. If two of his scales had been as long as the typical Thurstone scale, the reliabilities would have been higher. However, in the work of Rundquist & Sletto (**112**) it is found that the reliabilities for their scales of 22 statements (same length as Thurstone scales) are only .80 to .85.

Ferguson (**42**) has reversed the study of Likert, Roslow, and Murphy by starting with Rundquist-Sletto scales, which were constructed on the basis of the internal consistency criterion used by Likert. For four of these five scales, Ferguson found that the items were not well distributed over a continuum when scaled by the equal appearing intervals method; the items tended to pile up at either the favorable or the unfavorable end of the continuum. He found that the Rundquist-Sletto conservatism scale, the only one with scale values well enough distributed to permit the Thurstone type of scoring, yielded a correlation of .70 for the two scoring methods. From lack of well distributed scale values and this correlation Ferguson concluded that the Likert technique "does not obviate the need for" scaling by the Thurstone method. But this does not follow from the correlation, in the absence of any information regarding the homogeneity of the group upon which it is based and the reliability of the scores obtained by the median scale scoring; and it can be argued that it is unnecessary to have statements which yield well distributed scale values.

The writer is inclined to believe that some combination of these two competing techniques for scale construction would be better than either one alone. It would seem logical to expect that more reliable scales would result if the Likert method were modified to assure the selection of some items in the middle range of the favorable-unfavorable continuum, or if the equal appearing intervals technique, along with internal consistency, were used for item selection and the median check scoring were dropped in favor of the simpler scoring technique of Likert. Both methods have merits, and both have defects which might be overcome by a combination of the two. As indicated earlier, the Thurstone technique may not always lead to the selection of items which functionally fall along a single continuum, and it can be demonstrated algebraically that the internal consistency criterion alone is not a sufficient condition for selecting items which belong to a single dimension. Using both should help to realize this goal. The adoption of the simpler scoring scheme might allay some of the objections to the "scale" units involved in the Thurstone method. Some might be fearful that the neutral point be lost—this might be good riddance—but it is possible to specify neutral scores when the scoring is by the Likert scheme. A neutral score is earned by a person who tends to give an "undecided" response to statements even though they are piled up at the ends of the favorable-unfavorable continuum, or by a person who is totally inconsistent in his responses. Extreme scores are, of course, obtained on entirely different bases: in one case by the endorsement of one or very few of the statements at one extremity, in the other by strongly agreeing (disagreeing with items of reversed wording) with all or nearly all the statements which support one side of the issue. Yet both techniques yield scores which are distributed over many values, nearly always in a unimodal fashion, both skewed and symmetrical, with perhaps a greater tendency toward symmetry in the case of the Likert technique.

We will next consider briefly other methods, and modifications, which have been proposed for the measurement of attitudes. The attitude of a person may be rated by an individual well acquainted with him, or as a result of intensive interviewing. The latter will be discussed later in connection with problems of administration or data collection.

Self-rating has been used by Ewing (**35**) and by Riker (**104**). The former found rate-rerate (one week interval) reliabilities to be in the vicinity of .80, while the latter found that rate-rerate reliabilities for attitudes toward 13 different issues ranged from .50 to .86, and for intensity the values ranged from .66 to .99. This last figure is suspect. Riker found that, for attitudes toward the Negro, the Germans, the treatment of criminals, capital punishment, evolution, and communism, the correlation between self-ratings on an 11-point graphic scale and equal appearing intervals scales ranged from .55 to .84. If data were available for correcting these for unreliability, there is reason to believe that they would be raised to the high .90's. The correlations between the self-ratings for intensity and the Thurstone scale scores ranged from .57 to .75; corrected, these would likely approach .90, thus proving that the intensity factor is fairly well incorporated within the Thurstone scores. The self-ratings for favorableness and for intensity yielded correlations from .77 to .96, which may be spuriously high due to the possible correlation of errors. It would seem from Riker's results that the self-rating scheme has possibilities for securing a fairly satisfactory distribution of the attitudes of a group when a long scale is impractical. The question whether individuals are thinking of the same continuum, therefore rating themselves on a single dimension, would be difficult to answer.

The selection of items which yield differences between groups with known attitudes has been utilized by Zubin and Gristle (**133**) in constructing a test for measuring militarism-pacificism. The group taken as militaristic was an ROTC unit at the College of the City of New York, and the pacificists were members of peace organizations. It would seem that such a scheme should automatically lead to the inclusion of items which are valid, hence the validity of a scale so made up should be assured. There is no assurance, however, that items so selected will fall on a single militarism-pacificism dimension; the differentiation for an item might be due to a difference between the groups for a variable which is correlated with militarism-pacificism. A priori judgment in assembling items would help forestall such a happening. These investigators found that the split-half reliability for their 65-item test was .95. In this procedure involving the use of criterion groups, much depends on the real attitude of the individuals belonging to a group—why do they belong? For instance, at Stanford the use of the ROTC unit as a militaristic group might very well lead to the selection of items which aligned individuals along a continuum with pacificism (maybe) at one end and the "liking of horse back riding with the fair sex" at the other (supposedly militaristic) end.

In a study of the attitudes of labor union members toward a central theme, "the conception of workers as a potentially dominant class," Newcomb (**95**) discarded the Thurstone method as laborious, and inferior in reliability and in validity when scaled by a group different from the one with which it is to be used (evidently he was here ignoring the futile attempts of others to demonstrate that the scaling group's attitudes affect the obtained scale values). He also rejected the Likert technique as too demanding on the respondent with its consideration of both the statement and the five possible responses. Instead, he prepared 16 pairs of statements, each pair differing with respect to the central theme. The respondent was required to check the preferred item of a pair, and double check if strongly preferred. The writer is unable to see that choosing between two statements with two possibilities for checking preferences is any less demanding on the intellect than responding to statements of the Likert type. Furthermore, no criterion was used for eliminating faulty items, and there is nothing about the procedure which would indicate that a single continuum is involved. That this also results in inadequate reliability is shown by the fact that the median split-half reliability (stepped up) for 13 groups was only .81; for the heterogeneous combined group the reliability was higher, .91. The method used by Stagner (**118**) in 1936 resembles the Newcomb plan of item selection and elimination in that there was no plan. Thirty-five statements regarding Fascism were assembled as a test, the score being simply the number of pro-Fascist statements endorsed. The split-half reliability was .77 which is certainly below an acceptable standard, especially when the number of items is noted.

A technique which might overcome some of the invalidity of the verbal question-response type of test has been proposed by Proshansky (**98**). By an adaptation of one of the projection techniques, the Murray Thematic Apperception Test, he secured scores on attitude toward labor for two small groups which correlated .87 and .67 with the Newcomb questionnaire mentioned above. No reliability coefficients were reported, but from these two correlations, based on rather homogeneous groups, it can be inferred that the scheme yields fairly reliable scores. The possibilities of this type of attitude testing should be explored further. The above correlations do not, of course, prove the validity of either the projective or the Newcomb technique.

Rosander (**108**) has compared two scales, both constructed by the method of equal appearing intervals but differing in that the "behavior scale" contained verbally stated situations with instructions to endorse any of several reactions to each situation, while the "opinion scale" contained the usual type of statement. Both scales were designed to measure attitude toward "social equality of the Negro and the White." The reliability coefficients of the opinion scale were .72 and .90 and of the behavior situation scale, .74 to .87. The two types of scales correlated to the extent of .68, .81, .81, and .89, corrected for attenuation, thus indicating that they measure much the same thing. The author took these coefficients to indicate validity, but a far better indication of validity is the fact that northern and southern students were rather sharply differentiated by both scales.

Pace (97) has also used verbally stated situations calling for a verbal response as to which of several actions the respondent would take. The items were not scaled, but 10 judges ranked the stated actions as to social-political-economic liberalism. Arbitrary scoring weights were assigned to responses. As regards validity, the total scores differentiated completely (no overlap) between 25 known radicals and 25 known conservatives. Odd-even reliability coefficients for different groups varied from .50 to .92. Pace argued that ordinary scales have limitations as to validity, yet he went on to provide "further evidence of the validity" of his test by correlating it with the Rundquist-Sletto Economic Conservatism scale. The corrected coefficient was .72, which is rather high when it is recalled that Pace's test included social and political, as well as economic, conservatism. This type of scale is sufficiently promising to deserve further research.

The verbally stated hypothetical behavior situations with verbal responses can be contrasted with actual behavior in situations created for the purpose of ascertaining attitudes from what a person does instead of by what he says he would do, that is, his verbal responses to verbal statements. May and Hartshorne (83) used such a scheme for inferring attitude toward honesty.

A new approach to scaling is found in a 1944 paper by Guttman (52) who proposes a rational scheme, based on matrix algebra, for selecting items for scales to measure any type of psychological trait. The chief merit of the method, which is simple in application, is that it leads to elimination of items which are not on the principal continuum, thus assuring that a single dimension is involved in the retained items. Heretofore such assurance could be obtained only by factor analysis methods, which are so laborious as to discourage their use. Neither from the published paper nor through conversations with Dr. Guttman has the writer been able to see that this method adequately takes into consideration the ever present response errors or that ordinary standards of reliability can be attained by so few items as Guttman seems to think will provide an adequate scale. The troublesome validity problem is not solved by this "scalogram" technique, but its superiority on the single dimension problem, plus the safe prediction that reliability can be realized by taking sufficient items, makes this a most promising scaling method.

The reader will have noted that we have not had much to say about the validity of scores obtained by the use of scales. This void is mainly due to the fact that little has been done to establish validity. It is true that many comparisons have yielded differences in attitude scores in the direction expected on the basis of a priori judgment, but anything like a one-to-one correspondence between verbally expressed attitudes and actual behavior indices of attitude has not been found; in fact, seldom have investigators attempted to ascertain the *degree* of relationship existing between measures of attitude and criteria external to the scales. This is one place where more rigorous research is direly needed. The days of assumed validity should have long passed by.

There is ample evidence that scales can be developed which will yield satisfactory reliability, and it is possible to construct unidimensional scales. The writer sees no way of devising instruments which will yield units which are truly equal. Comparability of units can be attained by the use of standard scores or percentile ranks, perhaps by the method of equal appearing intervals, but such units permit rank ordering rather than true measurement. We would argue that we can secure a reliable and valid ordering of individuals on a single continuum, and that such can prove useful in the scientific study of the attitudes of men.

Very little has been said here concerning the wording of attitude statements. Aside from such obvious rules as those set forth by Wang (**125**), wording is not a major issue in the scaling problem whereas for the single question method much, as will be seen in the next section, revolves around the wording of questions and the type of response asked for. In constructing scales it is now generally, though not universally, agreed that it is better to provide for multiple responses rather than a dichotomy. The tasks set by scaling and by the single question technique are, of course, not the same. It is one thing to rank order individuals; the apparently easier task of assigning individuals to one of two classes is an entirely different matter.

III. SINGLE QUESTION OPINION GAUGING

The primary aim of opinion polling is to categorize individuals or their opinions into two or more classes which are conceived as either qualitatively or quantitatively different. Although this is rather inclusive, it is not our purpose here to discuss the problems which are peculiar to election straw polls. What we have chosen to call the fundamental requirements for an adequate attitude scale can be repeated here as the basic requisites for determining an opinion (or attitude) by the use of a single question. We refer to *reliability* and *validity*, both of which terms are used in the restricted sense, and to *dimensionality*. The application of these concepts in connection with polling techniques poses some problems which differ from those encountered in scaling.

In the case of an attitude scale, the *reliability* is easily calculated by correlating form versus form, or by the split-half method. To get at the reliability (accuracy) of responses to a single question, it would be necessary to ask the question in some alternate form or to reask the same question at a later time. In either case the time interval should be long enough that the respondents do not give the same responses because they remember their previous ones, but the elapsed time dare not be too long lest real, nonchance, changes take place. These obstacles can be overcome, but phrasing questions in alternate forms which are comparable is not an easy task. It will be noted that reliability of responses to single questions does not involve the notion of how accurately an individual is located on a continuum but how accurately he is classified into one of a few categories. Would he give the same response, hence receive the same classification, if he could be asked the same or a similar question under conditions such that he was amnesic for his original response?

In our earlier discussion of reliability it was pointed out that individual measurement is subject to two types of error, variable or random and constant or biasing. It was stated that biasing errors are of little consequence for scaled tests because when using such tests one is seldom interested in an elusive absolute value. Also in scaling, the use of several questions tends to average out the variable errors—increasing the length of a scale by the addition of items tends to increase its reliability. The situation, however, is entirely different when we look at the reliability of responses to a single question. The variable errors involved in categorizing an individual cannot be readily reduced by increasing the number of questions—the use of a battery runs right into the problem of scaling. It is true, of course, that the variable response errors to a single question will tend to be canceled in calculating the percentage of a group that holds a given opinion. The effect of these errors can be reduced still further by asking the group several questions on the issue, computing the separate percentages, then taking their average. This should yield a more dependable over-all percentage, but it would not provide a more reliable classification for an individual unless his answers were somehow combined, which again involves scaling. The problem of individual accuracy is of considerable consequence in any type of analysis whether it be for ascertaining the correlates of opinion or for the cross classifying of opinions. Obviously, if the obtained responses to a question were entirely chance determined, there would be no hope of finding correlates; and the extent to which chance errors are present limits the degree of correlation. In analyzing interrelationships, there is no use to handicap oneself unnecessarily with response errors, particularly with errors of unknown extension.

There are common sense ways by which single question reliability can be increased. Other things being equal, a question which is stated simply in easily understood words will tend to yield a reliable reply. Ambiguity, double negatives, and long questions tend to produce unreliable replies. Other things being equal, the greater the respondent's familiarity with the given issue, the more reliable the reply. The greater the personal relevance of the issue, the more dependable the response. The better the crystallization of opinion, the more reliably determined. It cannot be assumed that individual opinions about every issue can be reliably gauged but it is the definite responsibility of opinion pollers to determine *how* reliably individuals can be classified as to their opinions. This aspect of reliability has been practically ignored by the users of the single question technique, although they call their work "scientific."

The only information on the reliability of opinion questions is in Cantril's *Gauging Public Opinion* (18), and this pertains only to one question: "Do you think Roosevelt is doing a good job, only a fair job, or a bad job running the country?" For interview-reinterview, with a three weeks interval, the percentage of identical responses given by 286 persons was 79, which for the given situation tends to approximate a correlation coefficient of .90. This question is said to be one of the most stable, hence it may be inferred that other questions would yield lower reliabilities. By the same method 87% of the same group gave consistent answers to a question concerning whom they voted for in the last presidential election, and information given about car ownership (car or nor car) agreed only 86% of the time. Evidently the factual type of question is none too reliable, so one wonders just what the story is for opinion questions.

A paper by Hayes (56) throws some light on reliability via the consistency of responses to positively and negatively worded statements about the same issues. Regarding opinion concerning two issues, armaments and war debts, the degree of consistency as measured by the tetrachoric correlation coefficient varied from .60 to .70; regarding government ownership, taxes on risks, tariff, unemployment relief, and veterans relief the coefficients ranged from .40 to the .60's; for five other current issues the agreements ranged from the .30's down to .10. Strictly speaking, the use of positive and negative wording may not satisfy the requisite of comparable alternate forms, so the given coefficients may be underestimates of the reliabilities.

When we turn to the constant or biasing type of error, we find that the pollers have an appreciative awareness of the disrupting influence of bias, which is of paramount importance when the aim is to report an absolute—the percentage of a group who favor a certain action or who believe in something. To do this, it is necessary to avoid biasing errors, a subject to which we shall presently return.

The second basic requirement, that of *validity*, does not pose any problems unique to the single question technique. Does the question elicit a response which reflects the individual's opinion on the given issue, or is it a reflection of his opinion on some other issue? Is there any evidence that the verbally expressed opinion corresponds to a behavioral indication of opinion? Can a man's response be regarded as truthful? All this boils down to whether a person's opinion or attitude can be safely inferred from what he says. Or, to state the problem differently, it can be said that validity is a matter of interpretation. Suppose a soldier says he would like to go overseas soon, and that this is interpreted as indicating a desire to get into active combat. At once the question of validity faces us. Does his statement really indicate high fighting morale, or does it merely indicate a desire for change or for adventure? His verbal expression could be reliable in that he might consistently say the same thing, but to assume that it is a valid indication of a desire to fight may be far from the truth. Opinion pollers have done little to establish the validity of the questions they use. Here is a wide-open field for research, the difficulties of which should prove challenging to the best minds.

The few studies on validity have been concerned mainly with validity in a broad sense, e.g., polling results are checked against voting behavior for a group as a whole. That is, the check does not involve establishing the extent to which individuals voted in accordance with their responses to the poller. Agreement of over-all percentages is meat for the journalistic mill and perhaps sufficient for the crude comparison of groups, but it is difficult to see how a science can be built around opinions without first establishing individual validity.

Blankenship (**7**, **9**) has checked five types of questions, about horse racing and parimutuel betting, against voting intention, and found that the percentage of agreement ranged from 85 to 99. This is said to indicate validity but it is nothing more than agreement between two verbal expressions. One might have expected that Cantril's volume (**18**) would have tackled the problem of opinion question validity but such is not the case. Evidently "the serious problems encountered in every phase of the polling operation" (**18**, p. viii) do not include what the writer considers to be foremost in importance, namely, validity and reliability for the individual response. The volume does touch on some of the factors which contribute to unreliability and invalidity, but a frontal attack is sorely needed. In a 1945 article, Connelly (**24**) has characterized validity as the "real problem" facing opinion pollers. It is heartening that at least one poller has so expressed an awareness of this requisite.

That falsification in answering questions is indeed a fact is evident from the recent note by Hyman (**61**), who found that 17% of 243 individuals who were known to have redeemed war bonds answered that they had not cashed in any of their bonds. He also reports that the agreement between workers' statements and plant records on absenteeism ran from .88 down to .30, with a median tetrachoric correlation coefficient of .60 for 18 industrial plants.

Our third basic requirement for an adequate scale, *that it shall function in a uni-dimensional fashion*, must be stated somewhat differently as a requisite for a satisfactory single question. The categories into which individuals are classified on the basis of their responses to a single question can be either qualitatively or quantitatively different. The hope is that those who are placed in a class or subgroup are more alike in their opinions (or attitudes) than the total group. To argue that the classes represent quantitative differences is to assume differences in degree on, say, a favorable-unfavorable continuum. This assumption is doubtless true for a very large number of opinions, but in order to make a meaningful inference to the effect that those in one category are more favorable in their opinions, it is necessary to establish as fact that the responses given to the question do involve a single dimension. Otherwise, one might be dealing with an unknown mixture of quantitative and qualitative differences. If the purpose is to categorize individuals according to qualitative differences in opinion, then the greater the qualitative homogeneity of the opinions of those in a class, the more meaningful the classification. Thus one requirement for a good opinion question is that it shall lead to responses which either indicate differences in degree on a single continuum or permit classifying into qualitatively similar subgroups. This means that the question must be so simple that it will be interpreted in the same way by all respondents and that, in so far as possible, it should be answered from the same point of reference. That individuals may give similar responses but for different reasons is an obvious possibility. Intensive research is needed to determine the various reasons back of responses.

It is not here claimed that all the problems which beset opinion polling can be subsumed under the three basic requirements but these have implications bearing on practically everything that has been done and said about opinion gauging by means of single questions. A great deal of research has been directed at the *factors which tend to bias responses*. In fact so much is known about the variations which can be produced, and so little is known about which variation is most nearly correct, that one is apt to become pessimistic concerning the possibility of single question polling ever contributing scientifically useful data.

It is not our purpose to review or criticise the specific studies which have to do with the general dependability of results secured by the single question method. Rather it is our hope that a sketch of known and possible factors which affect the reliability, validity, and scientific worth of data based on responses to questions will aid in an evaluation of the entire polling procedure.

Of the several variations in ways of ascertaining opinion, the simplest is the direct question which calls for a "yes" or "no" answer. These responses are easy to code and tabulate, and their meaning is encumbered only by the meaning of the question. The assumption that a question has been interpreted in the same way by all respondents is made, usually without being stated explicitly, and is then forgotten, although it is amenable to at least partial check. The factors which may be operative in producing "yes" responses which are not of the same significance are many. The words in the question may carry *different connotations* for individuals of different cultural backgrounds or intellectual levels. For example, a surprising number of Negroes, interviewed orally, seemed opposed to government control of profits. Probing brought out the belief that God alone should exercise control of prophets! If a cross section of the adult population is to be polled, the vocabulary limitations of the subaverage must be kept in mind when questions are being phrased.

Aside from word meaning, the answer to even a simple question is partly dependent upon the respondent's *frame of reference* which may often differ from that hypothesized by the poller and from that of others who give the same response. The importance of the frame of reference is well illustrated by an unpublished check, by The Department of Agriculture's Program Surveys, on a recent Gallup poll. Gallup had reported that the majority of the people want things to remain the same, as judged by their responses to "After the war, would you like to see many changes or reforms made in the United States or would you rather have the country remain pretty much the way it was before the war?" Program Surveys asked the identical question, then probed further to learn how the respondents interpreted "changes or reforms." Sixty per cent had responded in terms of domestic (as intended by Gallup) changes and reforms, for 7% the frame of reference was not ascertainable, and the remaining 33% interpreted the question as referring to one of the following: technological improvements, changes in traditional political-economic structure, foreign affairs, immediate war conditions, personalized conditions, and desirable state of affairs. Of those who answered from the domestic point of view 60%, as compared with 46% of the total sample, voted for "changes and reforms."

Another frequently uncontrolled factor, which is closely related to frame of reference, pertains to the *understandability of the question and/or the issues involved*. For instance, how dependable an answer can one expect to the following question used by the Office of Public Opinion Research in a recent study (12): "For handling domestic problems like unemployment, the converting of war plants to peacetime use, or the demobilization of soldiers—do you think the Government should set up a central agency *now* with full authority to make plans and with full authority to carry out these plans as soon as the war is over?" The writer of that question certainly has an ivory tower notion regarding the intellectual span of the average adult. Questions are also being asked about the postwar world, the future peace, and other rather abstract issues, and the answers are being treated as though coming from a public with opinions well crystallized by experience and information. Such indeed become illustrations of "a consensus of worthless opinion is a worthless consensus of opinion."

Even if a simple yes-no question is clear, calls for the same frame of reference, and has a cognitive level low enough for all respondents, it entails other difficulties which can be subsumed under the *mechanics of questioning*. Some of the factors which may affect the answers are: positive versus negative statement of the question; attempted balancing of positive statement by ending with "or not"; loading by introducing emotionally charged words or phrases; the presence of contingent or conditional ideas; the influence of juxtaposed questions; suggestive elements; alternate wording; prestige elements; personalization of the question; stereotypes; technical words; biased wording; etc. Research directed at these factors indicates definitely that the percentages of yes (or no) responses can be varied. For example, Cantril (18) shows that "interventionist sentiment" between May and September 1941 showed an apparent variation from 78% down to 8% mainly as a function of the questions asked. Which is the correct percentage? No one knows.

Whether systematic investigation of such factors will ever lead to principles sufficiently general to permit their application to new situations is debatable when judged by the rather large amount of research already reported. Further research would serve a useful purpose if it led to more critical interpretation of over-all percentages, but such research cannot determine what particular combination of factors yields the "truth" unless external criteria are first established. Here we reach an impasse—seldom, except in the case of voting behavior, has an adequate criterion been found. It has been argued by some who are well aware of the dangers of taking percentages as absolutes that in comparing the relative standing of groups on an issue one need not worry about the influence of these factors. This assumes, of course, that any loading leads to the same amount of bias in the groups being compared.

Many objections have been raised against using questions which restrict responses to yes or no. Aside from the fact that the obtaining of information on certain issues may not always be amenable to yes-no questioning, it must be admitted that *qualitative and quantitative differences in opinion may exist among those who give a yes (or no) response.* In particular, there may be marked affective or intensity differences among those who give an affirmative answer; ditto, a negative response. In order to get at possible qualitative or quantitative divergencies among the yes (or no) responses, a number of response categories may be provided. Desirable as this may seem either from the respondent's viewpoint or as a scheme for securing added information, the inclusion of multiple response categories introduces a new set of methodological problems.

Possible variations in the response setup include the following: two, three, four, or more categories; extreme categories (effect of superlatives); omission or inclusion of a definitely labeled neutral response and varying forms therefor; unbalanced toward positive versus toward negative versus balanced categories; inserting an "undecided" response; reversing order of categories; including additional words in the possible responses; lengthy statements as responses; etc. Research has been directed toward discovering the effects of these and other factors at the response level. As in the case of question form, the results indicate that variations in the response structure lead to variations in classification and interpretation, but there would seem to be no absolute basis for determining the optimal response setup.

Other variations are being used. Sometimes the interviewee is asked to check an item or items in a list or to rank them in importance. The order of the items in the list may be a disturbing factor here. In some cases the respondent is asked to express his preference for a number of paired items. This, too, has its difficulties unless it is feasible to use a complete paired comparison scheme. At times, the so-called open-end or nondirective type of question is used, thus permitting a free or unrestricted response.

The open-end question coupled with highly trained interviewers has been advocated, particularly by Likert of Agriculture's Program Surveys, as one way of overcoming some of the difficulties arising from variation in question and answer form. The question is stated in the simplest possible manner and followed by further questioning to determine whether the respondent really understands the question and the issue, and to ascertain whether he maintains his opinion under mild cross examination. In other words, it is the job of the interviewer to be sure of the respondent's comprehension and to help him formulate, but not form, his opinion. Then on the basis of the interview, the respondent's opinion is categorized either by the interviewer or by an analyst or coder who studies the verbatim record.

The nondirective question, as frequently used in self-administered schedules, has the advantage that answers are not suggested by the listing of possible responses, and the disadvantage that the elicited response may represent what happened to be recalled or thought of at the moment. Difficulties are also encountered in coding or classifying the free responses. This type of question is most effective in the interview situation by virtue of the possibility of probing further by additional questions. Since these questions must be tailor made on the spot to fit the peculiar requirements of a given interview, it follows that such an interview cannot be standardized. Much is left to the ingenuity of the interviewer, and a high degree of training or an inborn knack becomes a prerequisite in order to objectify the interview situation. Interviewer bias is apt to be an uncontrolled factor.

If that bias can be overcome by training and proper supervision, and if the coding from the verbatim record of the interview can be objectified, it seems reasonable to believe that in general the use of nondirective questions will lead to more dependable results than those obtainable by restricted responses, especially when opinions on relatively complex issues are being sought. It is generally agreed that the open-end approach is of particular value during the preliminary phases of most studies, but there is a lively controversy, not entirely free of emotional cultism, regarding the proper place of free response interviewing as a method of ascertaining opinion. The extent to which the free and restricted response methods yield similar results is in need of further investigation. The supposed greater dependability of the open-end method must be weighed not only against certain subjective factors already mentioned but also against its much greater cost per case. It is more time consuming than ordinary interviewing, which in turn is more costly than self-administered schedules (not often feasible), and the work of the coder is slower. The coder also needs more training to overcome another source of bias and possible misjudgment.

The process is so slow that for certain governmental and journalistic purposes the reasons for making a given study may have vanished before the report is available, and so expensive that it is difficult to secure sufficient cases to permit breakdowns necessary for the analysis of relationships. Without doubt, however, the utilization of open-end questions by well trained and intelligent interviewers could serve as a useful check on responses obtained by more formal interviewing or by self-administered questionnaires. One would expect superiority for the open-end method when dealing with issues which for various reasons are not clear-cut, but for simpler issues this technique may not yield results any more dependable than those obtained by typical polling or questionnaire procedures. The devotees to interviewing by nondirective questions argue that this method can and should be used to determine the best combination of question form and response setup. (For a more detailed discussion of the open-end method, see references **73**, **80**, **117**.)

It takes no stretch of the imagination to see that by taking all the ways in which a question may be stated and all the possible variations in the response setup, it is possible by permutations and combinations to outline an enormous number of so-called research projects with the aim of studying the effects of purely mechanical aspects of asking questions and securing responses. A perusal of the literature and conversations with opinion researchers in government indicate that such mechanical research is regarded as of first importance in methodological studies. This may be due to the relative ease with which such research can be done and the fact that the investigator is fairly sure of obtaining positive results, in the sense of statistically significant differences. In the absence of any absolute or external criteria for judging which of many variations is best, and in view of the fact that opinion pollers are continuing to report percentages as though such figures had absolute meaning, it is difficult to see the value of this mechanical type of methodological study. Semantics and related disciplines may profit therefrom. Enough research along this line has already been done to indicate the fallibility of single question polling, and to provide caution signs for careful pollers, so it is doubtful whether more of this type of research is worth while.

Now that it is known that over-all percentages do vary with different ways of stating a question and with alternate response setups, what can be done about it? Obviously a percentage based on one form becomes a crude approximation. To know that another form will produce a lower or higher percentage does not constitute a safe basis for deciding the direction and extent of the possible bias produced by the particular form which has been used. In the absence of information concerning which of various alternatives involves the least bias, it might be assumed, perhaps falsely, that the least biased is that which gives a percentage near the average yielded by several forms. Preliminary work would be needed on each issue in order to learn which form was best by this criterion, and since the total possible number of forms is large, this task is scarcely feasible.

Another method which should yield about the same over-all percentage as obtained by this "average" question-answer form, and which rests on a similar untested assumption, is the split-ballot technique. This consists of preparing the schedule in several alternative forms and administering them so that the first interviewee answers on the first form, the second on the second, and so on through the sixth if there are six forms, then the seventh on the first, etc. For each question-answer form, some response is regarded as indicating, say, a favorable attitude; hence a count of the individuals, regardless of the form used, allows the computation of a percentage for the entire group. This over-all percentage will correspond exactly to the average percentage for the six forms if the number of cases (N) is the same for each form. Although a percentage arrived at by the split-ballot technique may be biased, it seems only natural to have more confidence therein than in the result from a single form. The public opinion pollers have, without doubt, ironed out some bias by this technique and thereby reported more dependable percentages.

That such a procedure does not end all the troubles which beset the pollers should be obvious. As already indicated, this type of average percentage may also be biased, and the inclusion of all possible forms of question and response setup, even if feasible, would not guarantee the elimination of bias. If breakdowns on such variables as age, sex, socioeconomic status, rural-urban residence, party affiliation, etc. are to be made, it is necessary to distribute the different forms evenly within each subgroup. This would not be impossible for a large polling agency, although if many forms were necessary, and such would seem to be the case, the difficulties would be great. All this pertains to reporting either one over-all figure, or percentages for various subgroups. Such may meet the requirements for reasonably dependable reporting for the press, for use by governmental agencies, and for the routine comparison of groups; but a science of public opinion will not make much progress without investigating the interrelationships and antecedents of opinions (or attitudes). Do those who have a given attitude on one issue tend to express certain opinions on other issues? What are the common psychological and sociological characteristics of those who are classified as having the same opinion on an issue? How do they differ from those who hold an opposite attitude?

To answer such questions it is necessary to make the breakdowns on the basis of the expressed opinions, but for this purpose the split-ballot technique is obviously inappropriate. In fact, its use for an issue upon which a breakdown is to be made would guarantee that those classified as, say, favorable on the issue would really be heterogeneous in their opinions thereon. For such analyses it is necessary to stick to one form of the question-answer setup, and naturally the one form to use is that which is least biasing, if this is known. And so we come right back to the fundamental problem of ascertaining how to select that question-answer form which avoids biasing the results. This, we repeat, cannot be done by demonstrating what seems obvious—that percentages can be made to vary by varying the form of the question and the permissible responses. An external criterion is absolutely necessary, but even if such were available one would be faced with additional complications. Suppose that various question-answer setups are to be tested against a known percentage, and it is found that a particular setup yields this percentage exactly. From this it would be concluded that a nonbiasing form has been found. What is the generality of such a conclusion? None. One wouldn't know whether this form would be best for another issue, nor whether agreement with the known percentage meant the absence of biasing for subgroups resulting from breakdowns by age, education, etc. The latter could presumably be checked in our hypothetical case. It would seem that the quest for an unbiased question-answer setup would be interminable.

Another problem which seems to perplex some critics of the polling techniques has to do with *intensity*. It is argued that it is not enough to know how many people favor a given proposition; we need to know something about the intensity of their attitude. Do all those who favor a thing do so with the same degree of affect? Are proponents or opponents of a social question more intense? The degree of emotion involved in the holding of an attitude or the expression of an opinion is of twofold significance: its bearing on action if action is possible, and on the stability of opinions. Presumedly, the greater the intensity the greater the likelihood of action, and the less the intensity the greater the possibility of changes in opinion.

Several schemes have been used to get at the intensity factor. Cantril and Rugg (21) have asked respondents to indicate whether their feelings were strong or mild, and Roper (**106**) indicates that he taps intensity by a scale of possible responses (strongly favor, favor, etc.). The most serious attempts to measure intensity, the studies reported by Cantril (**18**, pp. 51–65), will be reviewed briefly here.

One study involved two ratings of intensity, by self and interviewer, into one of five classes: strong approval, mild approval, no opinion, mild disapproval, strong disapproval, of Roosevelt's message to the dictators. It was found that both ratings were related to the respondents' opinions of Roosevelt as shown by their answers to other questions. The data do not permit a statement as to how high the relationships were. Next a graphic thermometer device for intensity self-rating was studied with reference to voting intention. The "very strong" supporters of a candidate differed from the "weak" supporters in their endorsement of statements regarding the candidate. Then a graded scale of only three statements, constructed by the Thurstone method, was utilized. The differentiability of this device proved to be so inadequate that the present reviewer completely rejects the statement that the "experiment demonstrates that simplified forms of intensity devices can be effectively used on public opinion polls" (**18**, p. 59).

Finally a far more extensive study is reported. This time seven devices were used, four of them under controlled conditions so as to determine their relative discriminating power. These four were a verbal self-rating of strength of feeling, self-rating on the graphic thermometer scale, a four-step logical or attitudinal scale, and verbal self-rating of certainty or sureness. The predictive efficiency of the four devices was to be determined by the extent to which intensity so measured on four key issues would predict the respondent's answer to certain test questions. Two of the key issues failed to function as expected, but the failure is rationalized, after the fact. For the two remaining key issues, percentage tables are presented for a total of five test questions, and from these tables each device is characterized as very good, good, fair, or poor. Since the criterion for this is not stated, and since one criterion which the present writer has applied to the tables fails to yield corroboration, the conclusions in Cantril's book are suspect.

Incidentally, this study is a good example of inadequate statistical treatment, and failure to report data in such form that any one can remedy the treatment. True, percentage tables are given (**18**, pp. 306–310) but the needed N's are missing, so our evaluation of the relative predictability of the four devices may be inadequate. At any rate, when one uses a simple criterion, the absolute differences between the percentages for agree versus disagree (or yes versus no) on the test questions, one finds, for example, that when the key question concerned trust in Russia, the two devices characterized as "very good" do not rank as high as the other two, one of which is said to be "good," the other "fair." For the key question on nationalizing industry our simple index places two devices labeled "very good" below a device labeled "good." We make no claims for such a simple index, but the reader of Cantril's book has a right to know the basis for the ratings given the four devices. To characterize predictions as very good without presenting a measure of association so that one can tell how good "very good" is injects an avoidable element of subjectivity into scientific research.

In this section we have attempted to point out some of the methodological difficulties associated with the study of opinion or attitude by way of the single question. We have found that investigators have practically ignored the problem of the reliability with which an individual's opinion is categorized. Nothing is known concerning the variable errors involved in opinion gauging. The question of error due to bias in question or answer form as affecting over-all percentages rather than individual responses has been studied by Blankenship (**6**, **7**, **9**), Cantril (**15**), Ghiselli (**50**), Hayes (**56**), Link (**79**), Matthews (**82**), Roper (**106**), Roslow, *et al.* (**109**), Rugg (**110**), Rugg and Cantril (**111**), and others. These studies merely demonstrate the important fact that variations in percentages can be produced by varying the question and/or answer form, but which setup is least subject to bias has not been ascertained. This ignorance concerning unreliability due to bias and to ordinary variable response errors, plus an equal ignorance of validity in the restricted sense, and scant knowledge as to qualitative or quantitative homogeneity of those assigned to an opinion category, to say nothing of the problem of being sure that issues are understood and opinions crystallized,—all these constitute a serious indictment of the claim that a science of public opinion is being or can be built around the single question procedure. The use of interview schedules with interlocking questions will not strengthen the edifice unless the responses to the questions are somehow combined into an index so as to obtain a more dependable basis for classifying individuals. Again the problem of scaling confronts us.

Not all of the difficulties in public opinion polling have been discussed here. Problems of interviewing or schedule administration, of sampling, and of analysis, have not yet been considered. Before going on to these, we shall make some suggestions as to what might be done about the grave inadequacies of single question polling.

I. *Reliability* (individual consistency or inaccuracy of classification due to variable errors) *can be readily studied, and therefore should be,* in order to learn how high, or low, it is and also which question-answer form is least subject to variable response errors. It is the writer's belief that the reliability for single questions will be found to be low; if measured by some coefficient equivalent to product moment correlation, we estimate that typical reliabilities will be in the .50's or .60's, sometimes lower and occasionally higher. Furthermore, it is unlikely that any amount of careful work will lead to an appreciable improvement in reliability over that attainable by using the question-response setup which turns out to be the most reliable.

II. Until rigorous research is done on validity, the pollers and the social scientists who use the single question method can expect continuing criticism regarding whether the questions used get at the opinion intended. *Validity studies can and should be undertaken.* Research on reliability and validity is, of course, humdrum work which is not productive of spectacular titles for articles and books, but some of the problems associated with determining reliability and validity should prove challenging to the most ingenious minds.

III. When one turns to the question of *whether one can be sure that responses are qualitatively or quantitatively comparable*, one is faced with a problem the solution of which has not, to the writer's knowledge, been found.

IV. The most difficult problem of all, however, has to do with *finding the form of question and form of response setup which is least subject to bias*. No more research is needed to show that the question-response setup is a source of possible error, the magnitude and direction of which cannot be established unless external criteria are available. This is seldom the case, so the bias in over-all percentages cannot often be determined. We have already indicated that polling has reached an impasse on this, and that the chase after absolute percentages is a delusion and a snare.

As the writer sees it, there are at least two ways in which social scientists can avoid this cul-de-sac. Both ways are relatively expensive, but fewer and better studies are perhaps desirable. The researcher might be forced to get along with fewer cases, and therefore be unable to make as many breakdowns—perhaps that would be a blessing too in that he would be forced to avoid some of the currently and routinely made breakdowns which seem to contribute so little to the understanding of opinions and attitudes.

The first way out is suggested with some trepidation as it has been the center of heated, though unpublished, controversy, which might have been avoided had the personalities involved been less uncompromising. This suggestion is *that greater use be made of the open-end, nondirective, intensive interview technique.* We have already discussed its merits and demerits. It is generally agreed that this method is of particular value during the preliminary or pretesting phase of most studies. It is the writer's belief, subject to correction by research, that the results obtained by highly trained interviewers using nondirective questions will possess greater reliability (accuracy for the individual), greater validity, greater homogeneity for those assigned to an opinion category, and less bias than results obtained by the ordinary polling procedure. If it can be shown that for a given issue the ordinary method yields data just as dependable, judged either by external criteria or by high correlation (say above .95) between the results obtained by the two methods, then the simpler procedure should be used. As to the intensity factor, one might expect the interviewer to be able to make an adequate appraisal thereof. It is not claimed here that the general dependability of the open-end method has been established—it just seems logical that it should be better than the regular procedure. Obviously, research is needed to delineate further the respective merits of the two approaches.

We have no hesitancy in making our second proposal, namely, *that single question opinion gauging be discarded in favor of opinion measurement by attitude scales*. By using scales a satisfactory degree of reliability (accuracy) can readily be attained. By factorial analysis or preferably by the Guttman (**52**) scaling method, uni-dimensionality can be guaranteed. There is evidence (**104**) that scaling automatically takes care of the intensity factor. Validity would have to be established but this is no more difficult for scales than for single questions, and with higher reliability the validity could be higher. Furthermore, validity is more meaningful when a single dimension is involved. The whole worrisome problem of bias due to question and answer form nearly vanishes. Scaling will be of no particular advantage to pollers and others who are satisfied with reporting percentages as absolutes. In fact, the placing of individuals along a continuum handicaps the percentage addicts. To use distributed scores as a basis for specifying what percentage of the population favor an issue is analogous to using a distribution of heights to say what proportion of people are tall. Scaling assumes quantitative differences—if qualitative differences are to be determined, scaling would be inapplicable.

There are additional advantages to measuring opinions by scales. By their use the social scientist is less apt to be misled or to mislead others with percentages. He has a more accurate instrument for classifying people into categories. He can, for example, segregate the upper, or lower, 30% with some assurance that the included individuals have been fairly reliably placed therein; by breakdowns on other variables he can make as meaningful analyses as are possible with the single yes-no question; he can use the distributed scores in his analytical work and thereby be able to express the closeness of relationships in an exact manner; and predictions can be more easily made from scaled opinions. If the aim is to analyze the variation in opinion on an issue, graduated scale scores greatly facilitate this; the variance can be broken down into component parts, a rather difficult if not impossible task when nothing but percentages are available. Furthermore, if one is interested in comparing groups obtained by different sorts of breakdowns, the analysis of variance technique can be used as the significance test (the requisite assumption of normality of distribution can be met by using transformed scores when necessary). Another advantage of scaled opinions is the fact that the variation within groups indicates the relative homogeneity of groups in their opinion about an issue.

If it is objected that a given problem does not justify the extra work entailed in scale construction and the added schedule administration time, there is reason to suspect that the project isn't of much value from a rigorous scientific viewpoint.

IV. ADMINISTRATION

We shall here consider briefly some of the problems associated with schedule administration, or the collection of attitude and opinion data. Certain aspects of administration are closely related to sampling, which will be discussed later. For the most part, data on opinions are secured by mail, by group testing, and by interviewing.

Mailed questionnaires have two major drawbacks, both of which have a direct bearing on sampling. First, it is difficult to build up mailing lists which are representative. The 1936 *Literary Digest* debacle will be recalled. Secondly, those who return questionnaires are apt to be atypical. Studies by Cahalan and Meier (**14**), Reuss (**103**), Shuttleworth (**116**), Stanton (**119**), and Suchman (**121**) have demonstrated that all too frequently the selective factors involved in mailed ballots and their return are apt to be highly significant. Mailed schedules or the telephone may satisfy the less scrupulous but such methods should be taboo.

The administration of schedules to groups, with each person checking or writing his own responses, is less expensive than interviewing. This is feasible for certain types of study, particularly when one is studying changes in attitudes or opinions produced by experimental procedures *and* when the necessity for securing a representative cross section of some defined population is not so important. For long schedules, self-filling of blanks by individuals in groups is less awkward than is individual interviewing. The group method has been widely used in studies based on college students; the sophomore in elementary psychology or sociology is easily accessible and usually willing to dawdle time on such schedules as are passed out to him. Since the college student is usually able to understand simple questions, there is no reason to suspect that he needs help in marking a questionnaire. As compared with interviewing, self-administration in groups rules out the possibility of interviewer biasing.

Even if it were possible to assemble groups for studying a cross section of the voting population, the write-in method would not always be advisable. Those at the lower educational levels might find the questions too difficult to answer without help. And a query can be raised regarding the comparability of results obtained by write-in and by personal interview. Some unpublished studies of the Research Branch of the Army's Special Services Division are of interest on this point. Near the beginning of the Branch's work the task of using a long questionnaire as a basis for interviewing a large number of troops looked impossible. Accordingly an experiment was set up for checking group versus interview secured data. Details cannot be given here, but there was evidence on the basis of percentages and scale scores that the two methods yielded comparable results except for those at the lowest educational levels. In the Army it is, of course, fairly easy to assemble groups and at the same time secure a good cross section.

For the great bulk of research in the field of public opinion, the chief method of securing opinion data is by face-to-face interviewing, and this seems likely to continue because of the near impossibility of bringing randomly selected individuals together. Even in the Army situation, interviewing was essential in the preliminary or pretesting stage of various studies. Quite frequently in both Army and civilian survey work it is necessary to begin with tentative questions on the issues to be studied, try them out by intensive interviewing of a relatively small number, then revise, eliminate, or add questions so as to secure those which seem adequate. Perhaps this process must be repeated several times before one hits upon questions which seem satisfactory. Sometimes the poller may be set the task of studying opinion about an issue which is none too well defined or about which so little is known regarding the reactions or thinking of people that it becomes necessary to use the open-end nondirective interview as a means of setting the questions. It goes without saying that this type of preliminary interviewing should be in the hands of trained and experienced individuals, and that much is gained by conferences between interviewers at this stage. No doubt some of the critics of the polling technique would be less skeptical about the dependability of results if they were required to participate in a few rounds of intensive pretesting. At least their subjective outlook on the whole procedure would be different.

The principal problems associated with interviewing have to do with the requisite training and with the possible influence of the interviewer on the respondent. This influence may result in bias or merely a reluctance of the respondent to express himself to a stranger. Those interested in a discussion of the sources of errors in interviewing as a method of collecting factual information will find the little book by T. E. Neely (94) worth while.

The type of individual used as an interviewer and the amount of training and supervision provided seem to vary from agency to agency. Apparently, Gallup depends mostly on part-time interviewers with a minimum of training and supervision. The writer has the impression that the National Opinion Research Center (Denver) and Roper's *Fortune* Poll pay more attention to selection, training, and supervision. Agriculture's Program Surveys, under Likert, employs full-time interviewers who undergo relatively more training and are rather carefully supervised. Little is known concerning the value of training and supervision. A reasonable hypothesis would be that more dependable results would come from those who are trained and supervised. The chapter on this subject in Cantril's book (**18**) presents so little evidence as to the difference between the training of the "trained" and "untrained" interviewers that one wonders why the study was ever made.

The extent to which respondents may tend to be influenced by having to give responses to a stranger has been studied by asking a random half of a sample to mark ballots secretly and drop them into a sealed or padlocked ballot box while the other half give responses to the interviewer on identical questions. There is some evidence (**3**) that more people give an undecided response, on candidates for election, in the interview situation than by secret ballot. The study by Field and Connelly (**47**) compared results secured by interviewing with those obtained in actual polling stations on election day. Again the secret ballot polled fewer undecided responses, this time on political issues. The responses secured by interviewing checked reasonably well with the "voting" results on two of the three issues, but on the third the difference in percentage of affirmative answers was too large to be chance. Cantril (**18**) found statistically significant differences on three questions, suggestive differences on four, and no differences on three questions for secret ballot versus interviewed groups. The differences seem to occur on issues on which a particular response carries social approval or prestige. It will be recalled that the experiment cited above, on self-administered schedules versus interviewing in the Army, did not produce response differences. This did not involve the secret ballot method, but schedules were unsigned and interviewee anonymity was guaranteed. Furthermore, it has been the subjective feeling of both civilian and soldier interviewers in the Army that the face-to-face interview situation does not inhibit the respondent. This may not hold for other groups.

A difference in status between the interviewer and interviewee is another source of error. Responses of Negroes to Negro and to white interviewers may differ (**18**) and the socioeconomic status of the interviewer has possible effects (**18, 65**).

The researches of Blankenship (**8**), Cantril (**18**), Katz (**65**), and Udow (**124**) seem to indicate that on certain issues the opinion of the interviewer tends to bias responses. Cantril concludes:

> Although interviewer bias exists, by and large the biases in one direction cancel those in the opposite direction, so that the over-all percentage of opinion is not likely to be significantly wrong. If an investigator wishes to minimize interviewer bias, he should choose an equal number of interviewers who are biased in different directions (**18**, p. 118).

So choosing the interviewers will scarcely be feasible for even a large agency when the schedule contains questions on more than a few issues. But suppose we stick to just one issue upon which opinion is fairly evenly divided, will the bias be averaged out by choosing an equal number of interviewers on each side? This depends upon the tenability of an assumption, namely, that the biasing effect of the pro-interviewers is the same in amount as that of the anti-interviewers. That such may be far from true is suggested by the fact that for many issues the supporters of one side are more apt to be evangelistic crusaders.

We are forced to conclude that the effect of the interviewer, whether unrelated to his own opinion (as in the secret ballot studies) or related to his own opinion (biasing), constitutes another source of error in over-all percentages. Also of importance is the fact that in the study of interrelationships we have another attenuating error. The possibility of overcoming this source of error by more carefully selecting and training interviewers should be investigated.

V. STATISTICAL ISSUES

Research on attitudes and opinions necessitates the use of certain statistical techniques as bases for making inferences. Usually the statistical treatment is relatively simple. The typical single question opinion study involves little more than the comparison of percentages for groups, while in the typical attitude study the means for groups are compared and correlations are determined. The statistical requirements are not much different from those in other areas of social science research. In general the value of a given study will depend in part upon adequate sampling, upon freedom from erroneous statistical manipulations, and upon the extent to which the statistical analysis and report are satisfactory. A thorough discussion of these statistical issues is neither feasible nor desirable here, but a brief sketch will serve a useful purpose if it leads only to an awareness of the statistical aspects of research in this area.

Sampling. All attitude and opinion studies are made on samples drawn from a universe or universes about which inferences are made on the basis of the statistical measures which describe the samples. In making such inferences, two types of error must be kept in mind—the random or chance sampling error and the biasing error due to possible selective factors. Actually, one finds here the usual sampling issues: the adequacy of the sampling techniques for yielding a representative cross section of a carefully defined population which itself is of sufficient generality to permit generalization of scientific value, and also the adequacy of sample size for reducing random errors to a desirable minimum for the total sample or for the subgroups involved in the analysis.

First, let us consider the sampling problems associated with the typical study in which attitudes are determined by the use of scales. Such studies are practically always made on college students. The real issue here is not only that of sampling adequacy but also of universe adequacy. Just what population is being sampled? The population of college students? All colleges? Just one college? Is this one institution typical or atypical? How is the sample drawn from among the students of the single institution? So as to get a cross section? Those who have even casually glanced at the literature will be able to answer these questions. Those who make the studies could also answer, but not without embarrassment. The fact is that all too frequently the investigator turns to the group that is most accessible. This is apt to be his own or a colleague's class of sophomores in either elementary psychology or sociology. If he is interested in comparing freshmen, sophomores, juniors, and seniors he can usually find an instructor who will willingly admit that the progress of science is more important than his Monday lecture. Little time is wasted in defining the universe and drawing a real sample therefrom, but one cannot infer from this that the investigator never heard of sampling. On the contrary, he demonstrates his competence by calculating sampling errors and talking about statistical significance.

It is, of course, obvious that this procedure does not involve the essential steps in sampling. The universe is seldom defined and the subjects are not often drawn in such a way as to constitute either a random or a stratified sample. The formulas used for computing the sampling errors are nearly always those based upon the assumption of random sampling, yet the two most elementary and fundamental principles of random sampling are not followed. That is, the sampling technique is seldom such that each individual in the universe has an equal chance of being drawn and that the drawings are independent of each other. So little is known about possible selective factors that it is even impossible to use the sample to define the undefined universe.

This inadequacy of sampling plus the fact that the groups used are from highly restricted universes leads one to question the value of a large proportion of the research on attitudes. Such generalizations as are drawn are so lacking in generality that it is difficult to see their worth as a part of social psychology as science. Sometimes an investigator is careful to circumscribe his conclusions with necessary qualifications as to their generality but that is no guarantee that these limitations will be heeded by others. Although it may be important to know something about the attitudes of the college sophomore, a science of attitudes cannot be built around this member of the species. We need more studies based on samplings from a far broader universe in order to have generalizations worthy of the name. Such studies will, of course, be more expensive to execute, but would it not be better to have one study with conclusions applicable to the generality than to have a dozen with conclusions inapplicable to the generality?

We hasten to add that the situation in the study of attitudes by scales is no worse than in other areas of psychology. The existing science of human behavior is largely the science of the behavior of sophomores. Too much research effort is expended on college students with subsequent waste of journal space devoted to speculation concerning whether the findings hold for mankind in general. There may be research problems in the attitude and opinion field which can be profitably attacked on the local campus, but it is the responsibility of the proponent of such research to show how the project can lead to generalizations which are not so restricted as to be meaningless.

When one turns to the typical single question or polling type of opinion study, one finds the attempt to sample from a universe broad enough to permit the establishment of generalizations which need not be qualified because of the possible peculiarities of the population being studied. The universe is usually defined as the population of voters or those of voting age. Sometimes a specified segment of the population is studied, not because of its availability, but because of interest in the group and a desire to compare it with the general population or with some other group. From the viewpoint of studying the opinions of adequate populations, it is fortunate that research on public opinion and its measurement got started alongside or with a polling agency which; in its efforts to predict the outcome of elections, was forced to consider a practically unrestricted population.

For nationwide or for regional surveys of either the total population or special groups, the problem of securing a representative, unbiased sample is being satisfactorily solved by the Bureau of the Census and by the Department of Agriculture's Program Surveys. Both these agencies are developing what is known as area or block sampling which must of course depend largely upon known characteristics of the total population. It has been shown that area sampling procedures will yield the correct proportional representation for the control variables, such as sex, age, rural versus urban residence, race, socioeconomic status, and geographic location, which are involved in the attempts at quota sampling by the private polling agencies. The recent paper by Hansen and Hauser (54) of the Bureau of the Census contains an excellent and convincing discussion of the advantages of area sampling.

The Research Branch of the Army's Morale Services Division had to face some peculiarly difficult problems in getting a cross section of the entire Army or, say, of infantrymen. It cannot be said that these problems were solved but, in view of practical considerations somewhat unique to the Army situation, a reasonably good sampling job was done. Sampling from the Army was apt to be disrupted by sudden shifts of personnel, but it had one distinct advantage over sampling from the civilian population: individual refusals were nonexistent, and the factors leading to missing an individual who had been designated for inclusion in the sample were not nearly as apt to be correlated with the variables being studied. Studies done in the Bureau of the Census, as reported by Hilgard and Payne (**57**), and Agriculture's Program Surveys (unpublished material) supply ample evidence that those who must be reached by two or more call-backs differ systematically from those found on the first call. That refusals may be a source of bias is shown by Cantril (**18**), who nevertheless argues that "refusals do not greatly affect the extent to which the sample secured by poll interviewers is a representative cross section of the population" (p. 122). This may be true, but when refusals run as high as 14%, it may well be questioned.

In the survey work done in the Bureau of the Census and in the Department of Agriculture, and to a certain extent in the Army the sampling is centrally controlled. Nothing is left to the judgment of interviewers, who seldom have an appreciation of sampling fallacies. This is in marked contrast with the procedure utilized by Gallup, Roper, and those making market surveys. So far as the writer can find, these gentlemen have never indicated precisely their sampling procedure. Apparently little central control is exercised except that the interviewers are scattered widely and are told to meet certain quotas as to age, sex, color, and economic status. It is reasonable to assume that the sex and color quotas are met by all except the possibly dishonest interviewer. When one considers the unreliability of the estimates of economic status, (interviewer agreement correlation of .63) reported by Cantril (**18**, p. 101), one is skeptical about the quotas for this control variable. The control of geographical and rural-urban quotas is entirely satisfactory as judged from figures given by Cantril (**18**, p. 145) for American Institute of Public Opinion (Gallup) and National Opinion Research Center (Denver) samples.

There is evidence, however, that the sampling procedures used by these two agencies lead to bias in the direction of too many from the higher occupational brackets. This type of selection is likely to be related to an educational bias. According to data given by Cantril (**18**, p. 148), both agencies sample about 20% too few cases from the grade school level, about 10% too many from the high school level, and 10% too many from those with education beyond high school. Cantril admits that central control of sampling is needed to overcome these selective factors, but he is inclined to believe that such would be "too complicated and costly for general use by polling organizations" (**18**, p. 148), which is another way of saying that quality of research is being sacrificed for quantity. Further evidence of educational bias is found in the Office of Public Opinion Research (Cantril) study on expected post-war migration, reported by Bruner (**11**). To find that 26% of workers in war plants have had some college education, as compared with the Census figure of about 10% and the Army figure of about 14%, suggests the operation of selective factors in the sampling procedures. We are told by Bruner that the interviewers secured workers for the sample from bus stations, parking lots, in lunch stands, and around pinball machines. It would be interesting to know just what factors do operate in the interviewer's choice of individuals for a sample.

A 1945 article by Cantril (**19**) is of interest in connection with the problem of sampling. He compared the responses obtained by different polls when identical questions were asked at nearly the same time by two or more of the following agencies: AIPO, OPOR, NORC, and *Fortune*. The respective N's were 3500, 1200, 2500, and 5000. The discrepancies between pairs of questions ranged from 0 to 12, with an average of 3.24 for 99 comparisons. Cantril hails this as "a highly creditable performance" considering the "expected margin of error" of "3 or 4 per cent." One notes immediately from the given N's that the discrepancies are larger than expected on the basis of mathematical error formula. What we have here are empirical discrepancies of the same order as those found in previous empirical studies, which, as will be seen in the next paragraph, led to the acceptance of 3 or 4 per cent as the "expected" margin of error. Even if the 99 differences were satisfactorily within the limits of predictable mathematical error, the result would not be infallible evidence that the polls were securing representative samples. All could be biased in the same direction—we have noted in the previous paragraph that the samples for three of these polls suffer from an educational bias.

In sampling for the purpose of predicting the outcome of elections one encounters grave difficulties. It may be relatively easy to secure an unbiased cross section of those who are eligible to vote but this might not lead to a good prediction as to what the subpopulation of actual voters will do. There are three possible sources of sampling error in straw polls: the ordinary random sampling errors, the biasing error due to selective factors in the picking of individuals by the interviewers, and the selective factors involved in voter turnout. The figures, presented by Gosnell (**51**), for the Gallup and Crossley 1936 polls would lead one to believe that the state by state error was too great to be attributable to random sampling errors. Likewise the analysis by Katz (**64**) of the 1940 election polls indicates that the regional errors for the *Fortune* poll and the state errors for Gallup's poll were greater than expected on the basis of chance. Gallup's national sample of 40,000 would mean that the standard error would be only .25%, hence his obtained error certainly (level of confidence represented by a P of less than .0001) should have been less than 1% rather than the obtained error of 3%. Without doubt bias of some sort was involved. That Gallup feels none too sure of the randomness of his straw samples is evidenced by the fact that his reported 4% margin of error is determined not from a sampling error formula but on the basis of empirical findings: it is, according to Katz (64), his average error over a period of years.

Katz (66) has also evaluated the performances of the 1944 election polls. Again the state by state errors of Gallup were so large and consistent as to suggest inadequate sampling. Crossley's poll suffered from the same limitations. Both these polls missed the national vote by a wider margin than did Cantril's OPOR, with a sample only one tenth as large, and the *Fortune* poll conducted by Roper. The uncanny accuracy of the *Fortune* poll in 1936, 1940, and 1944 is not evidence that Roper's sampling procedures are better than those utilized by the other polls. Rather it indicates his awareness of an educational or socioeconomic, hence Republican, bias which for these three consecutive forecasts has been balanced by not correcting for the relatively light voting in the South. Roper based his final prediction on the responses to a scale of attitude toward Roosevelt, mainly because he had found that the attitude scale tended to reduce the "undecided" percentage. That such a scale technique may be hazardous is indicated by Cantril's results, reported by Katz (66): an attitude toward Dewey scale gave 4.6% more Republican straw votes than did the Roosevelt scale. Cantril's forecast was obtained by averaging the results for the two scales. If Roper were aware of the possible loading in the scale technique, he must have reasoned that this would balance some other relevant factor. The unpublicised 1944 NORC poll erred in underestimating the national Democratic vote despite an effort to construct an accurate sample by the quota method. Again, it is likely that the educational or socioeconomic bias was operating.

There has been considerable alarm in certain quarters concerning the legitimacy of applying corrections to poll results. From the scientific point of view, there can be no possible objections to corrections for known bias in sampling or other faulty techniques *providing* the reasons for, and methods of, making the corrections are placed on record. This is not to say that corrections are to be preferred to the removal of their necessity. Corrections are apt to be too inexact, and at times involve guess work which may not be shrewd.

It should be emphasized that the failure of a polling agency to predict accurately the outcome of elections does not per se mean that its sampling techniques are inadequate for opinion studies. There are too many other factors involved; Katz (**64**) gives an excellent discussion of these.

Regarding the size of sample required for opinion and attitude research, it need only be said here that there are no simple, hard and fast rules. In general the nature of the problem and the type and number of breakdowns needed in the statistical analysis are the determining factors. Since the essence of a science is the establishing of relationships among its variables, breakdowns are necessary, and therefore the total N proposed for a study is not so important for judging the adequacy of the number of cases as are the N's in the required subgroups. Nonsensical as it may seem, one can find opinion studies by governmental agencies in which percentages are reported for subgroups of 50 or less; the writer has noted one percentage based on 14 cases! A consideration in deciding how many cases to use in correlational studies involving either single questions or attitude scales is the rather unstable nature of correlation coefficients. When subgroups are being compared or when interrelationships are being determined, it is well to remember that the acceptance of the null hypothesis is particularly fallacious if the number of cases is small. Differences of practical importance can too easily be brushed aside because of lack of statistical significance. For illustration, we cite the recent study of the 1942 elections by Cantril and Harding (**20**). It is reported that "an overwhelming proportion of the issues tested turned out to have little importance in determining the outcome of the elections." When 230 cases are broken down into six subgroups the sampling errors become so large that it is indeed difficult to reject the null hypothesis even though sizable differences exist between the population parameters.

When one is interested in the study of changes in opinion, the so-called panel technique is useful in reducing sampling errors for differences or changes. A representative group is set up as the "panel" with the idea that the members can be interviewed a number of times. The statistical advantage of the panel method is due to the subtractive correlational term in formulas for the sampling variance of differences. This technique does not, however, warrant the naive claim of Stonborough that "any change in a fixed panel is necessarily a real change and if the panel has been correctly set up to represent the universe under study, then a change in the panel *must* mean that a similar change in the universe has taken place" (120).

In concluding this brief discussion of sampling we would like to reiterate the following three points:

1. Too many attitude studies have been made on atypical groups. A real social psychology of attitudes is in need of research on groups of greater generality than college students.
2. The single question opinion studies by the several non-governmental polling agencies, AIPO, NORC, and OPOR, have been based on samples which are frequently biased. The dependence on interviewer selection of cases is likely to be the source of sampling bias.
3. Adequate and dependable cross sections can be obtained by area or block sampling.

It is not argued that all opinion and attitude studies should be based on samples of national scope. There are many projects which can be attacked locally with reasonable expectation that the findings will possess some generality. By "locally" we do not mean the local campus, but the local or some nearby community. Nearly every fair sized community will yield samples with sufficient variation on psychological and sociological characteristics to permit the determination of correlates of opinions and attitudes. The precise degree of such correlations may vary somewhat from community to community, but the divergencies will be minute compared with those between results for college students and for samples from any community which approximates typicality. There are, of course, many special groups within a community which might be profitably studied. Such community studies would involve field work and hence would be relatively more expensive than passing out questionnaires in a college class, but much could be gained by social psychologists from actual participation in field work.

Those who wish a technical discussion of sampling should turn to statistical treatises. For a good nontechnical discussion with special reference to opinion polling, the reader is referred to Stock (18, pp. 127–142), and to Wilks (129).

Erroneous Statistical Analysis. The purpose of this section is to point out some of the errors frequently found in statistical treatment, with the hope that such errors will not be repeated in future work. Since it is not desired to single out specific studies, references will be omitted but none of the illustrations given is fictitious.

One type of error has been found repeatedly in studies of changes whether produced experimentally or otherwise. In such studies it is considered good practice to have a pretest and posttest on the same sample of individuals. If scales are used, the change in attitude is measured by the difference between the mean pretest and mean posttest score; or if single questions are used, the change is the difference between percentages. In either case the sampling error for the difference must be determined by using the formula for the standard error of the difference which includes the correlational term. It is this term which is so frequently ignored. To argue that the omission of the *r*-term gives a conservative estimate of the sampling error is sheer nonsense. The fact is that omitting it gives a wrong estimate of error. Its omission springs from ignorance or carelessness, not from a desire to be conservative. It is true that if a difference is significant without the *r*-term it would be more significant with it. But why not use the term so as to be able not only to say that the difference is significant but also to say that it is significantly greater than some named value? Of greater consequence is the danger that the omission of this term may lead to acceptance of the null hypothesis when this hypothesis should definitely be rejected.

Another mistake is sometimes made in studying changes. An experimental and a control group are set up. If a change is significant for the experimental group but not for the control group, it is concluded that an effect has been demonstrated. The proper procedure here is to test the significance of the difference between the changes of the experimental and control groups, i.e., the net change. There are some studies in which the conclusions would be drastically altered if the net changes were evaluated.

A practice which can well be discarded is that of "explaining" a difference between means, or percentages, which is statistically insignificant. Similarly, a correlation which could easily be a chance variation from zero need not be explained or interpreted as indicating other than a lack of relationship. For example, in a study by a well-known social psychologist an *r* of $-.06$ was inflated into an italicised conclusion that a "direct affinity" existed between the two variables.

We have already mentioned as questionable the comparison of means based on different attitude scales.

A particularly fallacious procedure is the *blowing up* of a percentage. For example, in one study 35% of a sample of 400 said they would do a certain thing. This 35% applied to the total population of 20,000,000 led to the prediction that 7,000,000 people would so act. When one considers the sampling error of a percentage based on 400 cases, one readily sees that the correct figure could be anywhere between 6,000,000 and 8,000,000. To make matters worse there was no evidence that the sample of 400 was a good cross section of the 20,000,000.

We have already indicated that percentages based on rather small N's have been reported, and that accepting the null hypothesis is relatively more risky the smaller the N's. Despite the fact that the mathematical statisticians have provided ways for evaluating statistics based on small samples, the writer feels that the sampling techniques in social science are seldom if ever sufficiently adequate to justify using small samples. One wonders, for example, what importance can be attached to 1431 intercorrelations based on 31 cases.

The AIPO and NORC polls of opinion are based on stratified samplings, yet we cannot find evidence that sampling errors are evaluated by formulas appropriate to the stratified situation. This is discussed briefly in Cantril's book in a technical appendix by Mosteller (**18**, pp. 291–292), who concludes that "we may expect no great gain from the stratification over random sampling." If this is the case, it seems likely that the controls used in stratifying are not very effective. If at all effective, the sampling errors will be less than under conditions of random sampling, hence the use of formulas applicable to the latter will overestimate the sampling errors. This will not be a serious matter when judging the sampling stability of an over-all percentage. If a group difference is significant on the basis of the random sampling formula, it would be judged more significant if the formula for stratified sampling were used, hence no harm would result from using the wrong formula. Such is not the case when a group difference is on or near the borderline of significance. The use of the correct formula for the stratified situation might lead to the conclusion that the difference is definitely significant. How significant cannot be determined without using the correct sampling error formulas.

Occasionally opinion pollers state that for their N's a difference between percentages of, say, 5% is significant at the .01 probability level. If this holds for differences in the middle range of percentages, say 47% compared with 52%, it will not hold for differences at either end of the scale of percentages. Thus, for the same N's, the difference between 88% and 92% would not be judged significant on the basis of this rule-of-thumb procedure. Actually this four point difference would be significant at the .001 probability level. The danger in such an overestimate is as before: the null hypothesis may be accepted when it should be rejected. The nomograms of Wilks (**130**), reproduced in Cantril's book (**18**), for judging the significance of differences between percentages apparently do not allow for the fact that the sampling errors of percentages decrease with the distance from 50%. Wilks' chart of confidence limits for single percentages does take this into account.

Still other mistakes in statistical treatment have been noted but they are no more frequent in the attitude-opinion field than in other areas of social psychology.

Inadequate Statistical Treatment. The writer believes that the purpose of all statistical treatment should be to make analyses and to apply tests of significance which are as simple as possible yet suitable and adequate for the problems at hand. The researcher needs to keep in mind that statistical technique boils down to the simple task of calculating concise and convenient descriptive measures for masses of data and then drawing inferences therefrom. The descriptive terms include percentages, means and medians, standard deviations, measures of relationship or association, etc., for such groups as are involved in the breakdowns. Inferences are made about differences or relationships via appropriate sampling errors or significance tests. The reader of a report on an opinion or attitude study has the right to expect a full set of descriptive terms and specific indication of the significance tests applied.

A frequent shortcoming in the studies of attitudes by scale techniques is the omission of means, and more frequently, standard deviations. Such omissions make it difficult to compare one study with another. It is particularly disconcerting to have correlation coefficients reported without the accompanying standard deviations which are so necessary in interpreting correlations. Workers who are comparing various subgroups by attitude scales should examine the analysis of variance as a flexible technique for testing significance.

The rather persistent failure of the single question opinion researchers to give the number of cases in their over-all samples and in the subgroups involved in breakdowns does not inspire confidence in the scientific value of their data. Percentage tables or percentage bar charts are presented without any indication of the N's involved. The reader is helplessly and hopelessly dependent upon a verbal description which may or may not be buttressed with evidence as to what differences or relationships possess statistical significance, or as to how the significance was tested. In Cantril's recent book (18) there are approximately 400 sets of percentages without the basic N's. Yet this is a volume which according to its preface is "purely technical" and which "should discourage vague and unsupported criticism" of opinion polling!

In the *Public Opinion Quarterly* periodic reports of current findings of the several agencies are given in percentages for various breakdowns sans the N's so necessary for an evaluation of any hypothesis which might conceivably be checked by reference to these compilations.

In general, the statistical comparisons are made on the basis of standard errors of differences between percentages. This is adequate for a two by two table, but has some dangers for tables with more categories. The chi square technique is more generally applicable and possesses properties which make it a superior method for testing significance.

In studying relationships between opinions and other psychological or sociological variables, and between various opinions, the typical procedure is to set up contingency tables percentaged in one direction. This may seem plausible enough, but it should be recognized that the analysis of relationships via percentages is none too precise. Since practically all relationships so studied do and will show varying degrees of association, it would be far better to use, if possible, a measure which reflects the degree of correlation. That verbal description of percentage tables may be subjective and misleading is well illustrated by two examples culled from reports of a government polling agency. One relationship is described as "intimate;" had the analyst or reporter realized that the underlying correlation was only .60, it is extremely doubtful whether that misleading term would have been employed. The phrase "closely related" is used to describe a relationship which is a mere .40 on the product moment scale, and this same phrase is applied to various degrees of relationship. At another place a percentage table which leads to a correlation of .12 is referred to as showing a relationship. It is, of course, possible for such a correlation to be statistically significant if based on a sufficient number of cases, but its practical significance is nil. In regard to difficulties encountered in substituting verbal descriptions for more precise statements of correlation, it might be pointed out that statisticians have long stressed fallacies which are inherent in interpreting relationships from percentage tables.

In this sketchy discussion of statistical issues, we have attempted to show that far too many attitude studies have been based on atypical groups, such as college students; that the opinion studies have fared better in this respect by attempting, not entirely successfully, to use cross sections of the generality of adults; and that adequate samples can be obtained by the centrally controlled area or block sampling technique. We have pointed out some observed errors in the statistical analysis of opinion-attitude studies, the most frequent mistake being the use of an incomplete formula for the sampling error of changes. As to inadequate statistical treatment we have noted the failure to report basic descriptive measures, particularly the simple matter of N's in the opinion studies. We have suggested the use of two broadly applicable techniques for testing statistical significance: analysis of variance for studies based on scales and chi square for single question opinion research. It has also been suggested that in the latter type of study, greater precision can be obtained by substituting measures of association or correlation for percentage tables which purport to indicate relationships.

VI. Study of Changes

Changes in attitudes and opinions have been, and properly should continue to be, pertinent for those concerned with a science of attitude-opinion or with any aspect of social science wherein due consideration must be given to the influence of attitudes and opinions on the social behavior, actual or potential, of an individual or of groups. Any reasonable general hypothesis as to the origin or development of attitudes and opinions would have to place great stress on the role of learning, broadly defined as the modification of response tendencies by the impact of the entire culture which surrounds the individual during his life span. It would seem illogical to argue that attitudes and opinions are innately determined or that there are innate factors associated therewith, except that it might be hypothesized that certain very specific attitudes may depend upon so-called innate temperamental traits; that the intensity factor may be related to possible innate emotional tendencies; and that occasionally an attitude may be conditioned by one's physique. The chances for demonstrating that attitudes or opinions have innate correlates appear so small compared with the likelihood of finding cultural correlates, that it seems far wiser to concentrate research on the latter.

One method of investigating the influence of cultural factors on attitudes and opinions is through study of the conditions under which changes take place. The very fact that they do take place, whether accidentally or experimentally produced, can be taken as presumptive evidence that opinions and attitudes are learned. The problem is to determine the conditions under which changes occur or can be brought about. Ultimately this would involve a theory of learning, but whether a learning theory should precede or follow such inquiry is a debatable question. Those who are primarily interested in learning theory would argue that a consideration of such theory should lead to more fruitful hypotheses as to how changes can be produced, whereas those who are mainly concerned with attitudes could argue that they can induce changes which need to be explained by a learning theory. The writer prefers to avoid this issue.

There are, in general, two ways of studying attitude and opinion changes. One is by the use of trends based upon successive samples or cross sections from some defined population, and the other is by experiment. In typical trend studies group shifts are noted, whereas in experimental studies both individual and group changes can be observed. If significant changes take place in a well-controlled experimental study, the cause is known; but this is not always the case in trend studies—the investigator must rationalize shifts in terms of intervening events with no certainty as to which is causal.

Before the methodological issues involved in the study of changes are examined, a few general considerations should be discussed. When scales are used, the difference between means (or the mean of individual changes) can be taken as an indication of the shift which has taken place. Before one concludes that a real shift has occurred it is necessary to apply a significance test in order to be sure that the observed shift is nonchance. Statistical significance, however, does not prove that the change has social significance, and the magnitude of the critical ratio (difference divided by its standard error) does not necessarily reflect the amount of shift. Nor does the absolute difference between means indicate amount since the numerical difference is a function of an arbitrary measuring unit. Perhaps the best way to evaluate the amount of shift is to take it relative to the standard deviation of the first score distribution, or to consider the overlap between the first and second distributions. The thing to avoid is the frequent confusion between statistical and practical significance; the former can be objectified but the later cannot. For single question opinion gauging, the shift is in terms of a change in percentages and a nonchance change may or may not be of any social consequence. Suppose that the number of cases in two samples taken at different times were such that the shift from 38% to 42% for issue A and the shift from 48% to 52% for issue B were both statistically significant. The former shift might be judged of little social significance whereas the latter, being a shift to a majority, might be regarded as of great importance, provided one did not have too many reservations because of the factors discussed in Section III. As in the case of shifts in scores on scales, the critical ratio for percentage changes is not by itself an indication of the amount of change.

The evaluation of group changes is far simpler than the evaluation of individual changes. If scales are being used, a change for an individual can be appraised in terms of the error of measurement, but this cannot be done when the single question method is used. An individual shift on a scale can be judged as beyond the realm of chance response errors, therefore likely real; there is no way of being sure that an individual shift from yes to no, or vice versa, is not due to error except in the nonexistent case of perfect reliability.

It should be unnecessary to say that for the study of changes either by scales or single questions there are decided advantages in high reliability (accuracy). Furthermore, changes take on added significance if it is first demonstrated that the attitudes or opinions of individuals on a given issue have some stability in the sense of not fluctuating from day to day or week to week without apparent cause. To know just what issues people have relatively stable opinions about is also of interest.

Trend Studies. The study of changes by the trend method is best typified by the work of Cantril (**16, 17, 18**). Samples of the adult population are polled from time to time on the same question, and the resulting percentages are plotted as ordinates with time as the abscissa. For such ups and downs as occur an explanation is sought in terms of world or national events. If an event can be found which precedes a shift in public opinion and if it seems reasonable (without too much speculative argument) that the event would lead to the given shift, it may be rightly concluded that a causative factor has been found. It is obvious that such explanations cannot be as readily accepted as in a controlled experiment. *Ex post facto* rationalizations can sound convincing but be misleading. One fact which definitely emerges from the trend studies is that the prediction of public opinion for some future even though near date by extrapolation would indeed be hazardous.

Antedating the trend studies by polling were studies of attitudes based on successive college classes of, say, freshmen. A trend of attitude is noted and usually explained in terms of such large-scale events as the depression or the World War. Researches of this type are not as clear-cut from the sampling viewpoint as are the public opinion trend studies. In the latter the characteristics of the population are practically the same from time to time, whereas freshmen may differ from year to year. An observed "change" may reflect in toto or in part the changing characteristics of the freshman population rather than the influence of, say, the World War on attitudes.

The panel technique has already been discussed as a possible way of reducing random sampling errors in trend studies. In so far as the continued cooperation of the panel members can be obtained, and in so far as being repeatedly interviewed does not lead to changes, this method can be very sensitive for measuring shifts. The assets and liabilities of the method have been set forth in two papers by Lazarsfeld (**72**, **74**). One can readily agree to the following advantages: individuals improve in ability to express their opinions; additional personal information is securable each time; the individuals who change can be questioned in an effort to determine why they shifted; and the method is less expensive for some types of studies. Carried to the extreme of reinterviewing the same people over a period of 20 or 30 years, even if feasible, the panel technique would not necessarily lead to an accurate estimate of the general trend of public opinion, where "public" is defined broadly as the adult population. An upward or downward trend for the panel might be a reflection of the ageing of its members.

An extensive use of the panel method is reported in *The People's Choice* (**75**) by Lazarsfeld, Berelson and Gaudet. Out of 3,000 persons selected in Erie County, Ohio, by area sampling, four panels of 600 each were set up in such a way as to be comparable. The members of the main panel were interviewed monthly from May through November (post election) of 1940, while the other panels (controls) were interviewed in May and only once thereafter. Aside from determining much information about those who voted Republican as compared to those who voted Democratic, the reinterviewing of the main panel permitted not only a study of the characteristics of the changers but also an investigation of the factors associated with or leading to changes. This book is a "must" for all students of voting behavior and of methodology. Those who object to the rather small number of cases in some of the breakdowns will have noted one of the chief limitations of the technique: it is too expensive to permit adequate numbers for answering all the questions which tend to arise during the analysis of data.

The panel technique is very similar to that of testing and retesting college students as they progress from the freshman to the senior year. A number of researches have demonstrated that shifts in mean attitude scores for a variety of issues do take place as students go through college. Some importance can be attached to this knowledge, but the fundamental task is to pin down, if possible, the specific causes of the changes. It is not inconceivable that college attendance per se has little to do with the changes; a carefully matched control group of individuals not in college might show similar changes. Furthermore, the changes which take place in college are complicated by the selective factors involved in students dropping out of school. It is the writer's opinion that additional studies which purport to compare seniors with themselves as freshmen will contribute little to an understanding of the formation of attitudes and opinions.

Regardless of how small or how large the mean change in an attitude as students progress through college, one can be reasonably sure that individual shifts of considerable size have taken place. The correlations reported by Bugelski and Lester (13), Farnsworth (36), Hunter (60), and Newcomb (96) for attitude scores obtained as freshmen and as seniors are .42 or less, whereas for freshmen versus sophomore scores Corey (28) reports values ranging from .28 to .61 for various attitudes. Darley (30) found test-retest *r*'s in the .60's for autumn and winter versus spring testing on the Rundquist-Sletto scales. These correlations, particularly one of .13 for the Peterson war scale, seem to indicate a lack of stability for measured attitudes. Nothing is known about the individual fluctuation of opinions as gauged by the single question method. Research could be directed at finding the factors which lead to individual changes, providing the measuring were done by reliable scales so as to permit the singling out of individual shifts which are significantly larger than the error of measurement. The writer sees no way of doing this when the single question method is used. If 20% shift from yes to no or from one dichotomy to the other and it is known that the measurement or response error is such that a 10% shift can be expected, there is no way of knowing which of the 20% are real shifts and which are due to chance error.

The comparison of contemporary freshmen, sophomores, juniors, and seniors has been the basis for concluding that changes have taken place as a result of college attendance; but since it is possible, as shown by Newcomb (96), that the attitudes of these groups may differ at the time of entrance, such a conclusion is questionable.

Another type of trend study is that of tracing the "growth" of attitudes in children. This can be done by the cross-sectional method used by Horowitz (58), who "tested" school children at each grade level from the kindergarten through the eighth grade, or by the longitudinal or follow-up method. It is one thing to demonstrate that attitudes show growth, but quite a different task to account for the development. The fitting of curves, as in the Horowitz study, adds little to our understanding of attitude formation. In fact, curves and mathematical equations therefor tend to mask individual differences, particularly when no indication of variations about the curve is given, and it cannot be inferred that the growth of a given child will follow or parallel the curve based on averages unless it is established, by the longitudinal technique, that children maintain the same relative positions from year to year. It may be that the attitudes and opinions of elementary school pupils are too unstable to justify study at all, except perhaps to find out why an attitude "learned" today is unlearned next week. If particular attitudes and opinions do become relatively stable for an individual, it would be of interest to know what factors lead to the stabilization. If stabilization never occurs, the chase after causes may become a will-o'-the-wisp affair.

The study of trends in public opinion should contribute to an understanding of the factors which affect opinions, and should prove of value to the social scientist who is interested in social and political developments. The trend method does not admit the elements of control possible in the experimental approach, and furthermore the results are contingent on perspicacity in choosing the issues about which opinions are apt to be changed by events, usually unforeseeable. Yet in a sense, the use of trends is the use of the nation as a laboratory, hence no loss in reality is involved as is sometimes the case in experimental studies. That trend studies coupled with breakdowns can yield enlightening results is definitely suggested by Lazarsfeld's (71) study of changes in opinion during the debate and discussion concerning the appointment of Hugo Black to the Supreme Court. At first the line-up was political, later it was religious.

Changes by Experimentation. The literature contains some thirty reports of experimental studies of changes in attitudes or opinions. Most of these studies have involved measurement by scales. Some researchers have been primarily interested in attitudes per se and changes which can be produced, while others have been more concerned with the effectiveness of various media as possible molders of opinions or attitudes. Studies have been carried out on college, high school, and elementary school cases, hence the conclusions may not be applicable to other groups. One would expect the results obtained at the school level to hold generally for groups of similar level, whereas one cannot claim that findings based on college students can be generalized to other adult groups.

The experimental method has been utilized to investigate the effect of the following: propaganda, written and orally delivered in person or via radio; editorials and informative literature; classroom teaching of various sorts; group discussions; knowledge of expert or of majority opinion; and movies which might tend to shift attitudes. A little reflection will make it obvious that experiments based on different media cannot achieve the same degree of reality; if the subjects recognize the setup as artificial, the results may not be too dependable.

The essentials of a good experiment designed to determine the effect of a given experience on attitudes will be familiar to most readers although not all researchers using this approach have conformed thereto. An experimental group, for which an experience is provided, and a control group, which presumably is subject to the same influences as the experimental group except for the particular experience, are set up by random drawings of individuals from one defined population and by chance assignment to two groups. Greater precision can be attained by pairing of individuals for the groups, the matching to be done preferably on the basis of initial or pretest attitude scores or on the basis of control variables such as age, grade, sex, socioeconomic status, etc. The writer has discussed elsewhere the advantages of the matching technique (85). In practice it is not always feasible to use either the random or the matching method for forming groups. If one must resort to intact school groups, the aim is to choose two groups (each of which may be made up of subgroups) which are as similar as possible. Strictly speaking, this requires more complicated statistical treatment of results, but it is doubtful whether the use of the ordinary random error formulas would be seriously misleading.

Both the experimental and control groups are given a pretest, say on Form A; then after the experimental group has been exposed to the given stimulation, both groups are retested on form B. If only one form of a scale is available, or if the measurement is by single questions, the pre- and posttest must be based on the same form or the same questions. When the pretest and posttest data are in hand, the particular statistical treatment utilized varies. Suffice it to say here that it is inadequate and inefficient to test the significance of the difference between the two posttest means even though the difference between the two pretest means is also tested. Nor is it adequate to determine the significance of the difference between the pre- and posttest means (or the mean of the changes) for each group and conclude that an effect has been demonstrated if the experimental group shows a significant change and the control group does not. The proper thing to do is to test the significance of the net change, that is the difference between the changes for the two groups. This makes due allowance for any difference in the initial or pretest standing of the two groups and also brings out what is desired—the net difference in change reflects the influence of the experience interposed for the experimental group over and above the influence of other, and uncontrolled, factors common to the two groups during the interval. In computing the sampling error of the change for each group (needed in calculating the sampling error of the difference between changes for the two groups), one should avoid the frequent mistake of not including the correlational term.

If single questions are used, the changes will be in terms of percentages instead of means, but the general pattern of proper statistical treatment will be the same as just suggested.

There are times when a pretest is not at all feasible. This has been the case in some of the experimental work of the Research Branch of the Army's Morale Services Division on the efficacy of orientation films. An experimental group is shown a film, then tested at the same time that a carefully chosen control group is tested. Any significant difference between means (or percentages) for the two groups on this one testing (corresponds to the posttest) is taken as indicating that the film has had an effect. Obiously this design is not to be preferred over that permitting a pretest, but dependable results can nevertheless be secured providing the individuals can be selected either by strictly random methods or by pairing or matching in making up the two groups. If matched groups are used, the sampling error of the difference must include a correlational allowance for this fact. Some have been apprehensive about this single test experimental method because of the possibility of the groups not having the same initial standing. It should be noted, however, that the random formation of groups will not lead to initial differences larger than those attributable to chance, and that such chance differences will not produce nonchance differences in the end results. Pairing or matching of individuals on available control variables should help to allay apprehension concerning such an experimental design. Actually the situation is precisely the same as drawing samples, by the random or by the stratified method, from two populations; a population of those who can be provided with an experience like seeing a movie and a population of those who have not had this experience. The sample statistics are then used to determine the likelihood that the population parameters are different. The precision of the single test scheme will, in general, be less (sometimes much less) than that of the pretest-posttest setup, but this loss of precision can be overcome by taking more cases. For further discussion of these points the reader is referred to the previously cited article of the writer (85) wherein will also be found the appropriate statistical formulas for evaluating the difference between groups which have been formed either by individual pairing or by matching groups on control variables.

A quasi-experimental approach has been used in a number of attempts to demonstrate that changes in attitude and opinion occur as a result of taking certain courses of study. In the absence of a control group, one does not know whether the course or something else has produced such changes as occur.

In connection with the experimental production of changes the question is sometimes raised as to what types of persons show the shifts. This is a commendable quest but some who have taken it up have fallen astride the regression fallacy. If a group is given one form of an attitude test today and another form tomorrow, it will be found that those who were initially more favorable will be on the average somewhat less favorable on the second testing, whereas those who were more unfavorable initially will appear to become less unfavorable. Now, if in an experimental situation the mean for a group has shifted, say in the direction of a more favorable attitude, it will be found that those initially near average will have changed about as much as the total group, those initially favorable may have shifted very little or none at all, and those initially unfavorable will have changed much more in the direction of favorableness than has the group as a whole. This is the well-known regression phenomenon. The fallacy is that of concluding that the experience has had a differential influence according to an individual's initial standing. Actually this apparent differential effect may be nothing more than the regressive effect due to the ever present errors of measurement, which are, as we have seen, fairly sizable for attitude scales. The regressive effect of measurement errors will also tend to produce a correlation between initial standing and change. This is the same regression phenomenon which the writer has discussed in more detail elsewhere in connection with the study of IQ changes (84). It is not surprising to find instances of the regression fallacy in the attitude literature.

That attitude changes can be brought about by various means has been amply demonstrated by the experimental studies. It has also been shown that induced changes have persisted, though with some loss, for as long as six months. The writer is inclined to believe that, considering the possibilities for determining by the experimental method the formative causes of attitudes, only the surface has been scratched. Opinions could be studied likewise, but much more profitably if the single question were replaced by a scale. The possibilities are nearly unlimited, and the rewards in terms of understanding how attitudes and opinions come to be formed or changed are doubtless great. This is one way in which experiments can be set up with controls sufficiently adequate to guarantee the dependability of results, provided reliable and valid measures of socially important attitudes can be constructed.

VII. Correlations and Interrelationships

It can be argued that the fundamental task of any science is to discover and formulate the relationships between certain of its variables. The variates of a science may be thought of as those which are commonly accepted as such, plus those which have been newly isolated and those as yet undiscovered. Interrelations are sought as a part of scientific descriptions for the purpose of making predictions or providing "explanations." Thus the determiners of those variables known as attitudes and opinions are sought in terms of sociological, psychological, and sometimes physiological correlates; and the degree of intercorrelation among attitudes and opinions is of interest as a part of the description.

In the study of relationships two necessary considerations must be kept distinct. One pertains to statistical significance, the other to the degree of relationship. Before we can conclude that there is an association between two variables, some appropriate test of significance must be applied, but this is not sufficient even though a large number of researches on public opinion do stop at this point. A measure of the *degree* of association or correlation provides a more precise and objective description and permits a comparison of various relationships. High statistical significance does not necessarily mean a high relationship, nor does a low, though acceptable, level of significance always indicate a low degree of association.

Correlational analysis is most meaningful when classification on the variables being correlated is reliable. The attitude and opinion literature contains far too many instances of analyses based on measures of low reliability. In studying relationships it is also important to use groups for which the variation under consideration is unrestricted. If one wished to know the degree of correlation between height and weight for young adult males, one would not determine this on the football squad, yet such an absurdity occurs again and again in the study of correlates of attitudes. Just why one should allow a study of covariation to be limited by a restriction of the very variation being investigated is difficult to understand. This is not to say that research should be based on a group which is heterogeneous with respect to age, nor is it to imply that relationships based on groups so selected as to hold constant certain relevant variables are not of importance. The fact of the matter is that investigators using attitude scales have seldom chosen their groups with any serious thought as to the meaning of relationships which might be found. The public opinion pollers are not subject to this criticism.

Another question concerns what correlates and intercorrelates one should attempt to establish. One would naturally suppose that the answer to this would be found in the hypothesis stage of a study. Sometimes this is the case in attitude-opinion research, but more frequently the hypothesis step seems to have been displaced by a "let's see what this correlates with" or "wonder whether it will differentiate between this and that group." If anyone wishes to argue that such a criticism is unfair, let him then explain the abundance of negative findings reported in the literature. Surely carefully conceived hypotheses would not lead to such a dearth of positive results. For instance, why would any psychologist expect to find that attitudes of students toward Sunday observance, the church, and God would be related to receiving aid from the NYA? Or take a recent report (76) based on a mammoth questionnaire of 3000 items, laboriously filled in by 409 individuals. Some of the 3000 items were scored as a "conservatism-radicalism" scale, then it was ascertained that 1076 of 2600 items were related to conservatism-radicalism. Since by the statistical level of significance used, 5% or 130 of the 2600 items would be "significant" by chance, we infer that the author was correct in 946 and incorrect in 1654 of his hypotheses, that is if he started with hypotheses.

The Quest for Correlates. In the typical scale study the correlates of attitudes are determined either by breakdowns or, in case of graduated variables, by computing correlation coefficients whereas the single question opinion studies nearly always depend upon breakdowns. When breakdowns are used, the relationships are seldom stated very precisely.

In some scale studies, analyses are also carried out on the basis of the items. If the original scale is good enough to justify analysis of total scores, this item work is somewhat superfluous, especially when one considers the likely low reliability for responses to items. We have already stressed the importance of stating relationships in terms of degree of correlation. Certainly, if the task is to account for the variance in attitude by finding possible determiners among the so-called more fundamental variates, measures of correlation will be needed.

The search for the correlates of attitudes and opinions has been among both sociological and psychological variables. Typical breakdowns are for age, sex, rural versus urban residence, geographical region, racial or national origin, religious affiliation, education, socioeconomic or income status, political party, size of family, occupation, club memberships, etc. Percentages or means for the various subgroups are compared. It has been found that numerous attitudes and opinions do show differences for many of these breakdowns. Sometimes these differences are interpreted as throwing light on the origin of individual differences in attitudes or as indicating some of the determiners of opinions; at other times the emphasis is on the comparison of groups per se. Even if one were to grant that some combination of factors upon which the breakdowns are made could be utilized in "explaining" attitudes and opinions, it would be nearly impossible by this means of analysis to specify what proportion of the variation in attitudes had been so explained. But we can be reasonably sure that the residual variation left unexplained would be found upon further analysis to be surprisingly large, perhaps so large that it would have to be concluded that only a small percentage of the variance had been accounted for. The real challenge is the residual variation.

Correlates have been sought among other variables, such as specific information about issues, intelligence, personality characteristics, family and personal background factors, types of experience, etc. Parent-child correlations have been determined for some attitudes. Each attitude is apt to be related to a large number of interrelated and independent variables, including those mentioned in this and the last paragraph, none of which singly or in combination will be found to produce sizable correlations with the given attitude. It is the opinion of the writer that the quest for correlates among sociological and psychological variables, especially by the dragnet of many-item questionnaires, will continue to be rather unprofitable as a means for discovering the determiners of attitudes and opinions. Furthermore, attitudes and opinions which are relatively unstable are not apt to yield correlations with background factors. It may well be that some attitudes and opinions are too complex in origin to permit the finding of any very clear-cut determiners. If so, we may have to be content with the establishment of rough tendencies by the breakdown and correlational techniques. Such correlations as have been reported have been attenuated by measurement errors, by the use of selected groups with restricted variation, and possibly by the presence of more than one dimension in the measured attitudes and opinions.

Interrelationships. Those who have studied interrelationships among attitudes and/or opinions have been about as indifferent to hypotheses as those who have searched for correlates. In general, interrelations of attitudes have been determined on the basis of scales and on restricted, usually college, groups. The covariations are usually expressed in terms of correlation coefficients in studies based on scales, and as percentaged tables for cross tabulations in single question studies. The relatively low reliabilities for single questions become quite disturbing in determining the covariation of opinions. The reason for ascertaining correlations among attitudes is to learn what attitudes go together, or tend to form clusters. In the opinion gauging field, the term "patterns of opinion" has been used as somewhat analogous to clusters. Once a table of intercorrelations has been prepared, it is relatively easy to pick out the attitudes and opinions which tend to cluster together. This can be done roughly by inspection or more exactly by means of factor analysis. The patterns of opinion spoken of by Cantril (**18**, pp. 185–190), however, seem ill-defined and vague to the writer. Perhaps we have a bias in favor of correlation coefficients for this sort of thing.

The establishment and study of interrelationships of attitudes or opinions soon lead to two closely related questions: How many dimensions are needed to describe the attitudes and opinions of individuals? To what extent are attitudes general as opposed to specific?

The sense in which general and specific have been used in describing attitudes needs some clarification. Some have had in mind the connotation used by Spearman in connection with his two-factor theory of intelligence; but, instead of thinking of a general factor running through all attitudes, they have regarded a "general" attitude as restricted to a group of attitudes, e.g., conservatism-radicalism with respect to several social issues. Accordingly it would be postulated that an individual's score on a particular test depends in part on his general attitude and in part upon a factor unique or specific to the particular scale. It would be foolish to expect to find a general factor running through the aggregate of measured attitudes. This would imply that all attitudes are correlated, which certainly is not the case; hence the term "general" must be thought of as applicable to a family of attitudes such as conservatism-radicalism, morale, internationalism, etc. At this point we are not concerned with how many general or broad group factors might be required to explain the intercorrelations within a family of attitudes.

A second sense in which general and specific have been used in attitude research pertains to the form of statement utilized in attitude scales. Obviously a statement can be written in general form, e.g., "the Church is a valuable institution," or more specifically, "the Church is a solace to widows." The chief problem in this case is whether a scale made up of general statements and one made up of specific statements tend to tap the same attitude. The meaning of general and specific has, in this usage, a connotation entirely different from that of Spearman.

It is not our purpose to discuss the pros and cons of the controversy concerning whether attitudes are general or specific. The matter can best be settled by further research, and there is no need to hamper observation by postulating that personality is too integrated to permit specificity. Moreover the terms "generality" and "specificity" are relative to the situation—there are degrees of generality and degrees of specificity.

This whole issue, however, does have an important bearing on measurement theory. If someone proposes to measure a new attitude and assembles a set of statements for the purpose, he must show that the retained statements do tap something in common. The fundamental requisite of uni-dimensionality for attitude measurement means that the statements must have only one thing in common. It may be that general statements will be best for the purpose, but for some attitudes specifically phrased statements may be more adequate; even the general statements will always be found by statistical analysis to have some specificity. A second implication for measurement is that certain families of attitudes may not involve a general factor to the exclusion of group factors. The meaning of factors is never too clear even when they are derived by the neat mathematical devices known as factor analysis techniques. The general factor is frequently the most difficult to define—in fact, some factor analysts would not agree that such a factor exists. One school would cause it to vanish by rotations, another would keep it, but all agree that any of the factor methods will answer the question as to whether one or more than one factor is needed to explain the intercorrelations.

There is at least one family of attitudes whose meaning would be clarified if studied by factorial techniques. The attitude or attitudes in question have been the subject of a large number of publications, and many scales have been proposed as measuring instruments, but it is not always clear just what is being measured. Even though a given investigator knows what is involved in his measuring instrument, the attached label "conservatism-radicalism" is no guarantee that he has used the label in the same sense as that intended by any of his contemporaries. Not only do various scales constructed to measure "conservatism-radicalism" seem, by inspection, to be getting at different things, but even within a given scale there may be a mixture of kinds of "conservatism-radicalism." Furthermore, several of the Thurstone scales designed to measure rather specific attitudes are interpreted by some as measures of "conservatism-radicalism;" at times, four or five of the scales are used in a study to characterize people as conservative or as radical even though the intercorrelations among these scales may be low positive, zero, or low negative.

Apparently there may be different kinds of conservatism and radicalism, or this attitude may have many facets. Until such time as it has been established that radicalism-conservatism is a unity, it would be well for researchers to alter their labels to read "conservatism with respect to ——." The existing confusion might be alleviated by a thorough factor analysis of the several attitudes which might be considered as candidates for the conservatism-radicalism label. First, reliable scales should be constructed for each of the seemingly different aspects. Each of these scales should be as nearly as possible a pure measure of conservatism-radicalism on a given issue, i.e., no scale should be a conglomeration. Thus there would be assembled a battery of scales for measuring reactionism or conservatism versus liberalism or radicalism on a number of issues. These scales should be administered to a cross section of adults, with age controlled, for the purpose of calculating the intercorrelations. To insure some stability in the factor loadings, upwards of 400 cases should be used and the correlations should be computed by the product moment method. In spite of possible divergent interpretations of the factor analysis results, such a study would do much to clarify the meaning of this attitude cluster.

The judicious use of factor analysis might lead to enlightenment in other attitude and opinion areas. Little, however, would be gained by indiscriminate factor analyzing of every available table of intercorrelations, nor would a giant factor analysis involving all the available scales add much to our understanding of attitudes. It is very likely that many of the scales have little or nothing in common—factor analysis is designed to explain covariation, not its absence.

Unfortunately, the literature is already revealing some of the questionable benefits of factor analysis. To illustrate what this new statistical tool leads to when applied to the attitude field without sufficient foresight, we shall briefly review Ferguson's recent series of papers, the first published in 1939 and the seventh in 1944.

Ferguson began with the intercorrelations of the Thurstone scales for attitude toward: (1) God, (2) evolution, (3) birth control, (4) war, (5) capital punishment, (6) treatment of criminals, (7) patriotism, (8) communism, (9) law, and (10) censorship. In the first two factor analyses carried out (**39**, **41**), the last four scales did not behave in a rational manner, so they were purged. The remaining six scales yielded two factors: one was defined in terms of scales 1, 2, and 3 and named "religionism"; and the other, called "humanitarianism," involved scales 4, 5, and 6. Then in a later paper (**43**), by methods which would be questioned by expert factor analysts, Ferguson succeeded in "isolating" a third factor, defined by the last four scales and christened "nationalism." He has dubbed these three factors "the primary social attitudes."

One of his 1944 papers (**45**) reports a repeat analysis based on a new and larger sample of college students; since the same technique yielded the same factor alignment, we can at least be sure that "the primary social attitudes" are real in the sampling sense. The reality of these "primaries" is not readily grasped, however, when one considers the extent to which the defining scales are intercorrelated. For the first three, the basis for "religionism," the coefficients are .20, .15, and .46; for the three which are used to define "humanitarianism," .47, .24, and .39; for the last four, from which "nationalism" was elevated, .19, .24, .33, .32, .26, and .27. No further remarks are needed for those who have an appreciation of the fact that such low correlations indicate that the defining variables have little in common. Ferguson proceeded to set up scales for measuring the primaries (**40**, **43**. **45**), and for some of the measures he reports the impossible: higher "validity" than reliability. There is no evidence that the derived scales are unitary measures.

After "the primary social attitudes" had been discovered, isolated, and measured, Ferguson's next step was to search for their correlates (**46**); then—this is the capstone—he determined the correlations of "the primary social attitudes" with "national morale" (**44**). The lack of significant correlations in these two papers doubtless reflects a want of starting hypotheses.

Actually, granting that primary social attitudes exist, one would not attempt their discovery by way of a few selected attitudes. One would need to include *all* measured attitudes, with intercorrelations based on unselected cases rather than on college students. Even then there would emerge only the primary components underlying the attitudes for which there are measures, and these primaries might hold only for the population being sampled. Factor analysis may be a useful tool for certain problems, but one is skeptical as to its value in determining components in such a broad and diverse field as that of opinions and attitudes.

Perhaps progress in establishing the correlates of attitudes (opinions should be measured by attitude scales) can best be made by first ascertaining what attitude variables are judged by psychologists and sociologists to be of the greatest social and scientific importance. Then a concentrated effort should be made to find all possible correlates of each such attitude, or cluster of attitudes. Separate research programs could be set up for each attitude or attitude cluster. Each would involve an extensive program of research, but a lot of the hit-and-miss attempts could be avoided by giving due thought to the hypothesis stage. A large proportion of the false hypotheses might be eliminated in thorough critical discussion by a group of researchers with diverse theoretical backgrounds. If three or four such programs were carried out, we should be in a position to judge the fruitfulness of an extensive approach. It seems reasonable that a few such studies would contribute much more than the same amount of effort spread piecemeal over a variety of attitudes, some of which are of only passing or personal interest, and the study of which is too limited to provide anything more than qualified conclusions.

VIII. Studies of Morale

The national disunity in 1941 concerning our part in World War II, plus the continued lack of unanimity after we were plunged into war, was of such grave concern that there was rush to diagnose and prescribe a cure for the apparently low morale of the nation. Morale seminars were organized, governmental agencies called in social scientists to study civilian morale, and the War Department set up an organization to carry out research on the morale of troops. It would appear that the social psychologists of the country had only a few facts but much in the way of so-called principles to offer. Facts were lacking because so little direct research had been done on morale; nevertheless by 1942 a 463 page book on *Civilian Morale* (**126**) had been compiled by social psychologists. This purported to be scientific, but it is our opinion that the available factual data and legitimate speculations could have been summarized in less than 100 pages.

In the absence of a sufficient body of scientifically established facts and principles, the next best offerings of the social scientists were methods and tools for studying morale. The principal techniques available were those of attitude scaling and opinion polling. It is the purpose of this section to evaluate the chief efforts at defining and measuring morale. We shall proceed on the defensible assumption that scientific study of morale is impossible without some type of measuring stick; that is, we must be able to classify people as to their morale in order to discover what factors affect it.

What is perhaps the first serious attempt to measure and study morale is reported in the 1936 book by Rundquist and Sletto, *Personality and the Depression* (**112**). For these authors morale is "an exceedingly generalized trait," connoting "zeal, hope, confidence in oneself, and in what the future will bring . . . one's ability to cope with the future" (**112**, p. 201). Symptoms of low morale are: distrust of people, belief that the future is black, and that life is not worth living. They report a split-half reliability of .80, and test-retest (60 day interval) coefficients of .72 for men and .61 for women, for a scale of 20 items constructed and scored by the Likert technique. Evidence that the scale possesses some validity was obtained by comparing various groups. Their scale correlated with Hall's (**53**) occupational morale scale to the extent of .41 and .48. The Rundquist-Sletto scale has since been pitted by D. C. Miller (**87, 88, 89**) against the items of a 52 page questionnaire. This is another example of the dragnet method (cf. the 3000 item schedule mentioned in the preceding section).

Little need be said here concerning the Whisler and Remmers (**128**) three-way scale for measuring morale except that the reliabilities were only .55, .56, and .36, and that correlations with the Chant-Myers (**22**) optimism-pessimism scale were near zero. It is of some interest to note that the intercorrelations of the three parts of their scale were .57, .16, and .20. Whisler and Remmers' success in fitting perfectly a third degree parabola to four observation points does not, of course, add even an iota to scientific knowledge. This merely demonstrates a known mathematical fact in curve fitting.

The work on morale in industry also affords an indication of the measuring techniques available at the onset of the wartime crisis in morale. A typical procedure is that used by Kolstad (**69**) in a study of employee morale and attitude in a department store. A 10 item scale was developed for measuring that aspect of morale known as job satisfaction; then an attempt was made to find what specific situations or "gripes" were related to this general job satisfaction. The reader is referred to Blankenship (**5**) and Hull and Kolstad (**59**) for adequate discussions of methods of studying employee morale. For a good review of the morale literature which appeared prior to 1941 the reader should turn to Child (**23**). Here we shall consider briefly the studies initiated as a result of the war situation.

Miller (**90**) set himself the task of measuring "national morale," which he defines as pertaining "to all factors in the individual's life that bring about his energetic participation in the tasks which most effectively secure the national goals." He began by hypothesizing that national morale has five components, then deduced a total of 48 items from his hypotheses, and found that the 18 items having the highest internal consistencies included items deduced from all five postulated components. From this he concluded that his hypothesis concerning the components of national morale had been verified. Actually this does not constitute a verification, but does show that what one gets out of the statistical mill depends upon what was put into the hopper. Miller's resulting scale may have anywhere from one to perhaps six or seven components. The point is that he has not demonstrated by appropriate statistical treatment how many components are involved in his scale. Reliability coefficients running from .70 to .88 are reported in a second paper (**91**), which also gives a biserial correlation of .51 for low national morale and affiliation in organizations protesting the United States' entry into the war. Since this "validity" coefficient was based on a 2% dichotomy it must be regarded with some skepticism. Furthermore, certain statements in the scale are specifically concerned with interventionism, hence a relationship should not be surprising.

In this same paper, on the basis of differences found between samples from Washington State, Oklahoma A. and M., New Hampshire, Smith University (colored), and an unspecified college in Indiana, Miller concludes that there are regional differences in the national morale of college students. Maybe so, but no evidence is given that the five institutions are representative of their regions, nor that random samples were drawn from these five. The fact that he includes an item analysis comparison of the five groups indicates a lack of confidence in total scores or an admission that there may be 18 components in national morale. In a third paper (**92**) Miller reports a change to higher morale among Washington State students as a result of Pearl Harbor. The change is on the borderline of statistical significance. Again he turns to item analysis and discovers an upward shift for some items, a downward shift for others. This creates faith in neither the meaning of total scale scores nor what can be deduced from item analysis. But if one recalls the difficulties which plague the single question method, the inconsistencies for single items are not surprising. Diggs, Hanger, and Mull (**31**) have reported a correlation, based on college students, of .14 between the Miller and the Rundquist-Sletto scales. Evidently "national morale" is something different from morale as measured by the latter scale.

A different procedure has been followed by Harding (**55**) in constructing a morale scale. He started with 48 items based on a definition of morale similar to Miller's, found the intercorrelations among these items, and thence several clusters, four of which were used as the basis for a new edition of 59 items. These clusters have to do with confidence, tolerance, realism, and idealism. Then 20 of the 59 items were retained because they differentiated between a "high" morale group of 44 "distinguished social scientists and men in public life, each of whom is devoting a considerable part of his time to the furtherance of national morale," and a "low" morale group composed of 14 "surly" young men, mostly under 20 years of age, from a settlement house, and 32 men loafers on the Boston Common. That the "high" and "low" groups probably differed on a host of uncontrolled but relevant factors seems obvious. It is of interest to note that seven of the eight Rundquist-Sletto items tried on these two groups failed to show a difference. We infer from this that the Harding scale may have little in common with that of Rundquist and Sletto.

In an effort to find some of the personality-social correlates of morale, Sanford and Conrad (**113**) gave the Harding scale to 100 men and 173 women in a course in Mental Deficiency (what universe this sample represents is not specified). The report is based mainly on the results for men and presents the relationship of morale to 12 of 53 items in a questionnaire which the subjects also filled in. The authors do not reveal their reason for choosing to report on these particular items; if these 12 were the only ones which yielded relationships approaching statistical significance, we obviously have a hazardous capitalization on chance. No significance tests are reported; our casual calculations indicate borderline or less than borderline (5% level of confidence) significance. For example, "one of the most discriminating questions" concerned the number of children desired by the subjects. Calculation shows that the correlation between this and morale is .06 for 88 cases. By no stretch of the imagination or by no amount of statistical juggling can so small a correlation be considered significantly different from zero; yet these authors state that "it is clear that the high-morale men tend definitely to differ from the low-morale group in their expressed desire for children "(**113**, p. 13). This inflation of a correlation of .06 is accomplished by considering only the 9 high and 8 low morale men and ignoring the question of statistical significance. Sanford and Conrad found a correlation of .23 between the "confidence" subscale and the remainder of the Harding scale. Apparently the confidence aspect of morale has so little in common with the tolerance, realism, and idealism aspects that the meaning of a total score on the Harding scale is questionable.

A "life satisfaction index" of morale has been proposed by Goodwin Watson (**127**) This index is made up of four parts: liking of activities, liking of people, toughmindedness, and freedom from unhappy symptoms. These components intercorrelate to the extent of .54, .19 and four −.04's. The correlation of .54, between the two last named parts, could easily be spurious to the extent of .50 since toughmindedness is measured by only two items, one of which is nearly exactly repeated in the "freedom" component and the second of which is very similar to a "freedom" item. The "freedom" part contains eight items. Watson's index is based on markedly disparate parts, and as such is unequivocally meaningless.

Cronbach has defined morale as "the tendency of the individual to predict realistically the hardships or lack of hardships he will face in the future" (**29**, p. 12). He constructed a 50 item test for measuring "general optimism" and a 20 item test for "personal optimism," both having to do with the war. Apparently a number of items should have been eliminated from the scoring because of low internal consistency. The reliabilities were .78 and .56 for the general and personal scales for 980 high school boys; and .78 and .68 respectively for 1069 high school girls. The correlation between the two scales was .45 for boys and .40 for girls. A rough cluster analysis of some of the intercorrelations among the general optimism items led to the conclusion that morale, as here measured, is not a unitary variable, but may have three components: morale with respect to military progress of the war, civilian sacrifices and hardships, and personal freedom. A search for background correlates of morale proved futile.

On the assumption that one component of morale is "confidence or optimism," Conrad and Sanford (**25**) have prepared a 10 item scale to measure "optimism with regard to the military prospects for an early, easy victory," a 14 item scale to measure "optimism concerning the consequences of the war," and a 24 item scale "designed to measure general or personal optimism." To call these collections of items "scales" seems unwarranted since nothing is said concerning how the items were selected or whether any were eliminated. Apparently the old a priori method of the 1920's was used. Perhaps this partially accounts for the low form-versus-form reliability coefficients: .45, .62, and .60 respectively for the three scales and .62 for the first two scales combined. The corresponding split-half reliabilities were .49, .68, .70, and .65. A second article (**26**) was devoted to responses to the individual items even though it had been admitted that reliability of the single items is likely to be less than .25 (**25**, p. 302). These scales were given to University of California students, most of whom wore taking a summer (1942) course in Mental Deficiency. There are several points in this study which should be examined.

First, we note a tendency to assign absolute meaning to the mid-scale point as indicating a neutral position. Thus the group of students is characterized as being pessimistic about the consequences of the war and yet as possessing general and personal optimism. Perhaps this is true, but surely the authors are well enough versed in psychometrics to know that such absolutes are highly questionable. Changing the wording or form of the items can easily lead to a shift in means. A similar interpretation of arbitrary score units as having absolute meaning is found in their conclusion that the obtained standard deviations can be taken as evidence for lack of homogeneity or unanimity. In the article on individual items, it is again concluded that unanimity of response is lacking as judged by the spread of responses over as many as four of the five possible points for an item. The variance of this spread is about 1.00; now if we accept their own estimate of item reliability (or unreliability) as being less than .25, it can be said that at least 75%, perhaps 80%, of the lack of unanimity is due to response or measurement errors. If we reversed the Brown-Spearman formula, we would find .078 as the item reliability for the military optimism items and .115 for the "consequences" items. Maybe as much as 90% of the lack of unanimity is a function of response errors!

Some of the correlations reported by Conrad and Sanford are of interest. Military optimism and "consequences of the war" optimism correlate .33, or .57 corrected for attenuation, thus showing that the two scales have something in common. The correlation between military optimism and the combined scale (sum of the two) is .73; whereas "consequences" correlates .88 with the combined scales. The difference between these two coefficients was explained (**25**, p. 304) on the basis of the larger standard deviation for scores on "consequences." A logical and more pointed explanation is the difference in degree of spuriousness when scores on 10 items are correlated with scores based on these 10 plus 14 others, as compared with that when scores on 14 items are correlated with scores on these 14 plus 10 others. A correlation of .60 is reported for the combined "war optimism" scale with D. C. Miller's national morale scale. If this were corrected for attenuation it would become about .85, thus demonstrating that two of the many morale scales so far discussed really tap the same thing. What that thing is is not made clearer when one notes that an individual was considered to have low morale if he believed in 1942 that "Germany will probably bomb" our industrial centers in the East. Such an opinion would be tantamount to *high* morale in the sense of having confidence in leaders, since a desperate effort was being made to inculcate that very belief. One other finding of Conrad and Sanford should be noted: they estimate that the average intercorrelation among the military optimism items is .08, among the "consequences of the war" items, .12, and among the personal or general optimism items, .07. Evidently these aspects of morale have little in common, so the meaning of total scores is enigmatic.

Another major research on civilian morale appears as a 50 page report in Cantril's *Gauging Public Opinion* (**18**). This study reveals an awareness that morale is a complexity of attitudes. Twenty-five questions were carefully formulated to tap 12 components or aspects of morale deduced from a study of previous research. These questions were included in a ballot which was administered in November, 1941, by interviewers to a sample of 2543 adults, stratified for section of the country, city by size versus farm residence, economic status, sex, and age. It will be noted that this is the first published study of morale based on an adequate sample of the people. In order to determine whether the postulated components were independent of each other, the intercorrelations among 23 questions were computed, and the resulting table factor analyzed by the centroid method. Two factors were found sufficient to account for the correlations. The first centroid factor was identified as interventionism, i.e., high morale was associated with interventionism. If the reader is surprised at this, he should recall that what comes out of factor analysis depends upon what has been put into the mill.

The second centroid factor was called "degree of sophistication," but the opposite end of the axis seems to involve satisfaction with group progress and confidence in the leader, the news, and the armed forces. Apparently rotation of axes would not clarify the meaning of the two factors. Nor is the meaning of the second factor made more intelligible by the discussion on page 241 of "patterns" of intercorrelations. It is said that "The questions measuring *Awareness of the objective* and *Amount of information* are highly correlated with each other and moderately correlated with the Agreement and Determination questions." The last group of questions has to do with interventionism. What is meant by "highly" correlated? Coefficients of .30, .40, and .59. What is meant by "moderately" correlated? Eighteen correlations with a median value of .17, but including coefficients of .30, .33, .36, and .41. The above quotation is followed by "On the other hand, *Confidence in the leader*, *Confidence in the armed forces*, *Confidence in the news*, and *Satisfaction with group progress* are also highly correlated with each other and moderately correlated with Agreement and Determination." What is meant this time by "highly" correlated? Twenty-eight coefficients ranging from .66 down to −.13, with a median value of .24. And what does "moderately" correlated mean? Forty-eight coefficients running from −.18 up to .63 with a median value of .25.

The 19 questions yielding the highest loadings on the first factor were scored as a morale scale. Obviously such a scale is not unidimensional, but its saturation with interventionism is sufficient to produce differences along expected isolationist-interventionist lines. The scale reliability is not reported, but from the intercorrelations we infer that it would be between .80 and .85. With morale so closely identified with interventionism, it was of course necessary to revise the ballot for use after Pearl Harbor. The new ballot contained 27 questions designed to get at 16 postulated components of morale, and four questions designed to measure participation in war activities, "as a criterion of morale against which questions and components could be validated" (p. 243). This new ballot was administered to a cross section of 2539 adults in March, 1942. Most of the 16 components did differentiate between high and low participation individuals; we are not, however, provided with any numerical values by which to judge *how* important a component was. That the battery did not contain 16 statistically independent components is revealed by a factor analysis of the intercorrelations for the 27 questions. The three factors which accounted for the intercorrelations were named: reasoned determination to achieve the objective, confidence in leaders and satisfaction with traditional values. The importance of these factors can be judged by the average variance thus accounted for: 11%, 8%, and 6% respectively. The average uniqueness variance is 75% of the total. In reality, this means that the battery contains three group or common factors of minor importance and 27 specific factors. For only one question is the communality greater than the uniqueness. For half the questions, the communality is less than .25, which is what one would expect when the average of the absolute values of the intercorrelations is only .14. This leads one to suspect the meaning of "generalized" attitudes as discussed on page 272. It should be borne in mind, however, that one reason for the lack of any marked general factor or factors in such a battery is the undoubtedly low reliabilities for the responses being correlated.

The most recent article on morale measurement is that of Estes and Estes (**34**) who propose "miniature" scales, of from 3 to 8 questions each, for measuring 11 aspects of morale. The test-retest reliabilities range from .51 to .91, median of .82, hence it is doubtful whether the "scales are sufficiently reliable for use in tracing changes in attitudes." One certainly would be skeptical of individual changes. The total score based on all 47 items yielded a reliability of .93, but the meaning of total scores is obscured by the very low interrelationships among the 11 aspects—for two samples, only 6 of the 55 intercorrelations are consistently greater than .30. The authors state that all the aggressive attitudes tend to go together, but examination of their data indicates that this holds for only one of their samples.

What conclusions can be drawn from this brief review of the attempts to measure and analyze morale? It seems obvious that morale is not an entity, that there are many "morales," that there has been an open season for calling a wide variety of things by the name of morale, that the several aspects of morale can be fairly reliably measured by scale techniques, that nothing is known about the reliability of single question indicators of morale, that the instruments for gauging morale pertaining to war effort are soon outdated, that some types of morale are mutually correlated though in varying degrees, that little is to be gained by factor analyzing intercorrelations based on single and likely unreliable questions, and that much of the available research on morale suffers from a lack of hard-boiled critical mindedness.

It may be that "morale is a lot of little things," hence difficult to define as a scientific concept. It seems more reasonable, however, to believe that these "little things" are not entirely independent, either statistically or functionally, that certain of them tend to go together or form clusters, and that such clusters are conceivably independent of each other. Validation of scales designed to measure these several factors would require as many external criteria as there are factors. Two jobs need to be done: *the determination of the dimensions of morale and the construction and validation of scales for measuring these dimensions.*

Where to begin is a difficult question, but the following procedure is suggested. Start with postulated components. Here we need to make a distinction between two concepts. Although "component" and "factor" are frequently used synonymously, we shall for the purpose of this discussion use the term "factor" as in current factor analysis, i.e., "factors" are statistically independent in the sense that they are the orthogonal axes or dimensions necessary to explain a set of intercorrelations. If certain oblique axes are needed, then certain factors would not be uncorrelated. By "components" we shall mean those aspects which are logically postulated as the ingredients of morale. It is not implied that these postulated components are mutually independent. They constitute the many and varied "morales" which have already been defined in the literature and any additional concepts which workers in this area care to propose. Each postulated component should be carefully defined. Perhaps it would be necessary to call in an expert in semantics in order to determine whether certain definitions were really different. There is no way of knowing how many components might be postulated. If an unwieldly number, say more than 60, were proposed, it would be advisable to submit the list to a group of competent judges to see if some might be eliminated. It is certainly not unreasonable to require that a proposal should have objective acceptance beyond the subjective horizon of its proponent.

The next step would be to construct a uni-dimensional scale for measuring each of the several components. Use of the Guttman (52) technique would greatly facilitate the meeting of this requirement. Scales are suggested rather than single questions so as to have reliable measures. A Guttman type scale for each component would also avoid the absurdity of having several questions which supposedly tap a given component turn out, as sometimes happened in the Cantril study (18), to have little or nothing in common.

When reliable uni-dimensional scales have been constructed, they should be administered to a cross section of the adult population. A carefully stratified sample of 400 cases would be sufficient for determining the intercorrelations of the several scales. It would be more economical to compute the correlations by the product moment method than to have to double the size of the sample in order to have the same degree of stability for tetrachoric coefficients. The table of correlations among the postulated components would then be factor analyzed to ascertain how many factors underlie the several aspects of morale. If the centroid method were used, rotations would be necessary and there would be some difficulty in finding appropriate descriptive labels for the factors. Then scales could be constructed for measuring each of these "primary" factors, validity studies could be made, and finally a search for psychosocial correlates of the "morales" could be undertaken.

This may sound like an ambitious program but it would be one way to bring some order into a rather chaotic field of research.

IX. Concluding Remarks

An attempt has been made herein to provide a critical appraisal of what the writer considers to be the chief methodological problems involved in research on attitudes and opinions. It has been assumed that progress in developing a science of attitudes and opinions, or the fruitful scientific study thereof as parts of whatever discipline, will continue to be largely dependent upon the success with which attitude and opinion variables are measured. In this connection, we have stressed the basic need for reliability, validity, and uni-dimensionality for the instruments or devices used to classify or measure individuals with respect to their opinions or attitudes. Attitude scales can be so constructed as to attain satisfactory reliability. Unitary scales can be developed by the Guttman scaling technique. Validity can be established, but not without more effort than is usually expended.

Opinion gaugers have paid scant attention to these three basic requirements. They have, however, demonstrated that the general dependability of the single question method is none too high—a variety of percentage variations can be produced by changing the wording and form of the question and by altering the response setup. Enough is now known concerning the possible variations associated with question-answer form to justify the conclusion that the typical poll results are subject to a variety of possible errors. Because of its many grave difficulties, we have suggested that the single question technique be replaced by attitude scales. Has a science ever suffered because of refinements in measuring techniques? It is true, of course, that less research would be possible if scales were substituted for single questions, but one good study might be worth several based on inadequate techniques.

The statistical issues in attitude-opinion research are not different from those encountered in other fields of social science. Inadequate analyses and statistical errors have been plentiful, but as more statistical sophistication is acquired, one can expect adequate statistical treatment with fewer errors. Securing representative samples from the general population is no longer an impossible task, but the study of attitudes has been so limited by the campus-bound inertia of researchers that generalizations have tended to hold only for college students, rather than for man in general.

A study of changes under experimental conditions with necessary control groups would seem to be a fruitful way of ascertaining some of the factors which contribute to the formation or learning of attitudes and opinions. Lifelike experiences can be provided in order to see whether changes take place. The other principal method for finding attitude and opinion determiners is to look for correlates. This search could be facilitated by a greater stress on the critical formulation of hypotheses.

We have suggested that the judicious use of factor analysis might help to bring some order out of the chaos in certain areas, particularly that of morale and of liberalism-conservatism.

This appraisal has not touched on certain fields of application of attitude-opinion techniques, e.g., market and consumer research, which the reader will find discussed in Blankenship's *Consumer and Opinion Research* (**10**). Nor have we been concerned with all the problems associated with election polling. The reader is referred to Gallup and Rae's *Pulse of Democracy* (**49**) for greater detail on this topic. The serious student will not be aided by Gallup's recent publication, *A Guide to Public Opinion Polls* (**48**), whereas he will definitely benefit from a study of the analyses by Katz (**64, 66**) of the 1940 and 1944 polls.

We have not found an integrated summary of the findings of public opinion research. Newcomb (**93**) has provided an excellent summary and synthesis of attitude studies up to 1937, with a good discussion of methodological issues.

It will have been noted that very little has been said herein concerning the work on opinions and attitudes by the various governmental agencies, such as Program Surveys (under Rensis Likert) in the Bureau of Agricultural Economics, OWI's Surveys Division (Keith Kane, Rensis Likert, Julian Woodward, Elmo Wilson, Daniel Katz, NORC, *et al.*), and the Research Branch of the Army's Morale Services Division (Samuel Stouffer, Leonard Cottrell, Carl Hovland, *et al.*). Certain individuals in these agencies have given such loud acclaim to their interest in and work on methodological problems that some social scientists have been quite optimistic about the contributions to techniques which might emerge from this extensive work. Can it be said that these agencies have developed and perfected techniques which will aid in the solution of problems in the attitude-opinion field?

The writer had close contact, from 1941 through the winter of 1944, as a consultant to one of these agencies (the Army's) and in order to have a broader basis for answering the above question he has conferred with key men in the other agencies. All were very cooperative and willingly talked over their methodological problems, but a documented answer to our question is impossible because so little concerning the work on methodological issues can be found in the records of the agencies —the available information seems to be cached in the minds of staff members. It is apparent, therefore, that any appraisal must be tentative; yet the writer feels reasonably sure that the following observation is irrefutable: no marked contributions to methodology were made during the first 30 months of the war effort.

Minor advances have been made. The Census Bureau and Likert's group in the Department of Agriculture have, as previously noted, succeeded in ironing out the biasing bugs of sampling, but this was being done before the United States entered the war. Likert has made progress in refining open-end question interviewing. Guttman's scaling method originated in connection with his work as a consultant to the Research Branch of the Army's Morale Services Division. When completed, this method will be superior to the equal appearing intervals and internal consistency techniques, and will be applicable to other types of psychological measurement. All the major agencies have carried out studies of a methodological nature, but the striking thing is the fewness and limited nature of such studies. One looks in vain for work on the fundamental problems of reliability and validity. No doubt many individuals in these agencies have acquired a wealth of information and experience which may later lead to major improvements along methodological lines.

The above remarks are not to be interpreted as an indictment—perhaps those who profess their interest in methodology have been unable to do much because of time pressure. Even those who have hoped that their work will result in residual contributions to scientific techniques have found that their active interest in studying methods tends to wax and wane. Temporary spurts of enthusiasm were often dwarfed by the immediate problems of the day, and sometimes dampened by budgetary difficulties. A reading of the recent paper by Woodward (**131**) makes it quite evident that those in government who were doing opinion research during the war were definitely handicapped in various ways. The writer hopes that he is in error in failing to see any reason for optimism concerning war-born advances in the attitude-opinion field of research.

BIBLIOGRAPHY

1. ALLPORT F. H., & HARTMAN, D. A. The measurement of atypical opinion in a certain group. *Amer. polit. Sci. Rev.*, 1925, **19**, 735–760.
2. BALLIN, M. R., & FARNSWORTH, P. R. A graphic rating method for determining the scale values of statements in measuring social attitudes. *J. soc. Psychol.*, 1941, **13**, 323–327.

3. Benson, L. E. Studies in secret-ballot technique. *Publ. Opin. Quart.*, 1941, 5, 79–82.
4. Beyle, H. C. A scale for the measurement of attitude toward candidates for elective governmental office. *Amer. polit. Sci. Rev.*, 1932, **26**, 527–544.
5. Blankenship, A. B. Methods of measuring industrial morale. In G. W. Hartmann & T. Newcomb (Eds.), *Industrial conflict; a psychological interpretation.* New York: Cordon, 1939. Pp. 299–312.
6. Blankenship, A. B. The choice of words in poll questions. *Sociol. soc. Res.*, 1940, **25**, 12–18.
7. Blankenship, A. B. Does the question form influence public opinion poll results? *J. appl. Psychol.*, 1940, **24**, 27–30.
8. Blankenship, A. B. The effect of the interviewer upon the response in a public opinion poll. *J. consult. Psychol.*, 1940, **4**, 134–136.
9. Blankenship, A. B. The influence of the question form upon the response in a public opinion poll. *Psychol. Rec.*, 1940, **3**, 345–422.
10. Blankenship, A. B. *Consumer and opinion research.* New York: Harper, 1943.
11. Bruner, J. S. How much post-war migration? *Amer. J. Sociol.*, 1943, **49**, 39–45.
12. Bruner, J. S. Public thinking on post-war problems. Washington: Nat'l Planning Ass., Planning Pamphlet No. 23, 1943. Pp. 36.
13. Bugelski, R., & Lester, O. P. Changes in attitudes in a group of college students during their college course and after graduation. *J. soc. Psychol.*, 1940, **12**, 319–332.
14. Cahalan, D., & Meier, N. C. The validity of mail ballot polls. *Psychol. Rec.*, 1939, **3**, 2–11.
15. Cantril, H. Experiments in the wording of questions. *Publ. Opin. Quart.*, 1940, **4**, 330–332.
16. Cantril, H. America faces the war: a study in public opinion. *Publ. Opin. Quart.*, 1940, **4**, 387–407.
17. Cantril, H. Public opinion in flux. *Ann. Amer. Acdd. polit. soc. Sci.*, 1942, **220**, 136–152.
18. Cantril, H. *Gauging public opinion.* Princeton: Princeton Univ. Press, 1944.
19. Cantril, H. Do different polls get the same results? *Publ. Opin. Quart.*, 1945, **9**, 61–69.
20. Cantril, H., & Harding, J. The 1942 elections: a case study in political psychology. *Publ. Opin. Quart.*, 1943, **7**, 222–241.
21. Cantril, H., & Rugg, D. Looking forward to peace. *Publ. Opin. Quart.*, 1940, **4**, 119–121.
22. Chant, S. N. F., & Myers, C. R. An approach to the measurement of mental health. *Amer. J. Orthopsychiat.*, 1936, **6**, 134–140.
23. Child, I. L. Morale: a bibliographical review. *Psychol. Bull.*, 1941, **38**, 393–420.
24. Connelly, G. M. Now let's look at the real problem: validity. *Publ. Opin. Quart.*, 1945, **9**, 51–60.
25. Conrad, H. S., & Sanford, R. N. Scales for the measurement of war-optimism: I. Military optimism; II. Optimism on consequences of the war. *J. Psychol.*, 1943, **16**, 285–311.
26. Conrad, H. S., & Sanford, R. N. Some specific war-attitudes of college students. *J. Psychol.*, 1944, **17**, 153–186.
27. Corey, S. M. Professed attitudes and actual behavior. *J. educ. Psychol.*, 1937, **28**, 271–280.
28. Corey, S. M. Changes in the opinions of female students after one year at a university. *J. soc. Psychol.*, 1940, **11**, 341–351.
29. Cronbach, L. J. *Exploring the wartime morale of high-school youth.* Stanford Univ.: Stanford Univ. Press, 1943.

30. Darley, J. G. Changes in measured attitudes and adjustments. *J. soc. Psychol.*, 1938, **9**, 189–199.

31. Diggs, E., Hanger, E., & Mull, H. K. Morale in the college situation in relation to the morale scale of Rundquist and Sletto. *Amer. J. Psychol.*, 1942, **55**, 561–562.

32. Dudycha, G. J. A critical examination of the measurement of attitude toward war. *J. soc. Psychol.*, 1943, **18**, 383–392.

33. Ericksen, S. C. A skeptical note on the use of attitude scales toward war: I. In 1940, 1941. *J. soc. Psychol.*, 1942, **16**, 229–242.

34. Estes, W. K., & Estes, K. W. Minnesota studies in war psychology: I. A set of miniature scales for the measurement of attitudes related to morale. *J. soc. Psychol.*, 1944, **20**, 265–276.

35. Ewing, T. N. A study of certain factors involved in changes of opinion. *J. soc. Psychol.*, 1942, **16**, 63–88.

36. Farnsworth, P. R. Changes in "attitude toward war" during the college years. *J. soc. Psychol.*, 1937, **8**, 274–279.

37. Farnsworth, P. R. Shifts in the values of opinion items. *J. Psychol.*, 1943, **16**, 125–128.

38. Ferguson, L. W. An item analysis of Peterson's "war" scale. *Psychol. Bull.*, 1938, **35**, 521.

39. Ferguson, L. W. Primary social attitudes. *J. Psychol.*, 1939, **8**, 217–223.

40. Ferguson, L. W. The measurement of primary social attitudes. *J. Psychol.*, 1940, **10**, 199–205.

41. Ferguson, L. W. The stability of the primary social attitudes: I. Religionism and humanitarianism. *J. Psychol.*, 1941, **12**, 283–288.

42. Ferguson, L. W. A study of the Likert technique of attitude scale construction. *J. soc. Psychol.*, 1941, **13**, 51–57.

43. Ferguson, L. W. The isolation and measurement of nationalism. *J. soc. Psychol.*, 1942, **16**, 215–228.

44. Ferguson, L. W. The relation of the primary social attitude variables to "national morale." *Amer. sociol. Rev.*, 1944, **9**, 194.

45. Ferguson, L. W. A revision of the primary social attitude scale. *J. Psychol.*, 1944, **17**, 229–241.

46. Ferguson, L. W. Socio-psychological correlates of the primary attitude scales: I. Religionism; II. Humanitarianism. *J. soc. Psychol.*, 1944, **19**, 81–98.

47. Field, H. H., & Connelly, G. M. Testing polls in official election booths. *Publ. Opin. Quart.*, 1942, **6**, 610–616.

48. Gallup, G. *A guide to public opinion polls.* Princeton: Princeton Univ. Press, 1944.

49. Gallup, G., & Rae, S. F. *The pulse of democracy.* New York: Simon & Schuster, 1940.

50. Ghiselli, E. E. All or none *versus* graded response questionnaires. *J. appl. Psychol.*, 1939, **23**, 405–413.

51. Gosnell, H. F. How accurate were the polls? *Publ. Opin. Quart.*, 1937, **1**, No. 1, 97–104.

52. Guttman, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, **9**, 139–150.

53. Hall, O. M. Attitudes and unemployment. A comparison of the opinions and attitudes of employed and unemployed men. *Arch. Psychol., N. Y.*, 1934, **25**, No. 165, 1–65.

54. Hansen, M. H., & Hauser, P. M. Area sampling—some principles of sample design. *Publ. Opin. Quart.*, 1945, **9**, 183–193.

55. Harding, J. A scale for measuring civilian morale. *J. Psychol.*, 1941, **12**, 101–110.

56. Hayes, S. P., Jr. The inter-relations of political attitudes: II. Consistency in voters' attitudes. III. General factors in political issues. IV.

Political attitudes and party regularity. *J. soc. Psychol.*, 1939, **10**, 359–398; 503–552.

57. HILGARD, E. R., & PAYNE, S. L. Those not at home: riddle for pollsters. *Publ. Opin. Quart.*, 1944, **8**, 254–261.
58. HOROWITZ, E. L. The development of attitude toward the Negro. *Arch. Psychol., N. Y.*, 1936, **28**, No. 194, 1–47.
59. HULL, R. L., & KOLSTAD, A. Morale on the job. In G. Watson (Ed.), *Civilian Morale.* Boston: Houghton Mifflin, 1942. Pp. 349–364.
60. HUNTER, E. C. Changes in general attitudes of women students during four years in college. *J. soc. Psychol.*, 1942, **16**, 243–257.
61. HYMAN, H. Do they tell the truth? *Publ. Opin. Quart.*, 1944, **8**, 557–559.
62. JOHNSON, H. M. Pseudo-mathematics in the mental and social sciences. *Amer. J. Psychol.*, 1936, **48**, 342–351.
63. JONES, V. Attitudes of college students and the changes in such attitudes during four years in college. *J. educ. Psychol.*, 1938, **29**, 14–25; 114–134.
64. KATZ, D. The public opinion polls and the 1940 election. *Publ. Opin. Quart.*, 1941, **5**, 52–78.
65. KATZ, D. Do interviewers bias poll results? *Publ. Opin. Quart.*, 1942, **6**, 248–268.
66. KATZ, D. The polls and the 1944 election. *Publ. Opin. Quart.*, 1944, **8**, 468–482.
67. KIRKPATRICK, C. Assumptions and methods in attitude measurement. *Amer. sociol. Rev.*, 1936, **1**, 75–88.
68. KIRKPATRICK, C., & STONE, S. Attitude measurement and the comparison of generations. *J. appl. Psychol.*, 1935, **19**, 564–582.
69. KOLSTAD, A. Employee attitudes in a department store. *J. appl. Psychol.*, 1938, **22**, 470–479.
70. LA PIERE, R. T. Attitudes vs. actions. *Social Forces*, 1934, **13**, 230–237.
71. LAZARSFELD, P. F. The change of opinion during a political discussion. *J. appl. Psychol.*, 1939, **23**, 131–147.
72. LAZARSFELD, P. F. Panel studies. *Publ. Opin. Quart.*, 1940, **4**, 122–128.
73. LAZARSFELD, P. F. The controversy over detailed interviews—an offer for negotiation. *Publ. Opin. Quart.*, 1944, **8**, 38–60.
74. LAZARSFELD, P., & FISKE, M. The "panel" as a new tool for measuring opinion. *Publ. Opin. Quart.*, 1938, **2**, 596–612.
75. LAZARSFELD, P. F., BERELSON, B., & GAUDET, H. *The people's choice.* New York: Duell, Sloan, and Pearce, 1944.
76. LENTZ, T. F. Personage admiration and other correlates of conservatism-radicalism. *J. soc. Psychol.*, 1939, **10**, 81–93.
77. LIKERT R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.*, 1932, **22**, No. 140, 1–55.
78. LIKERT, R., ROSLOW, S., & MURPHY, G. A simple and reliable method of scoring the Thurstone attitude scales. *J. soc. Psychol.*, 1934, **5**, 228–238.
79. LINK, H. C. The eighth nation-wide social experimental survey. *J. appl. Psychol.*, 1943, **27**, 1–11.
80. LINK, H. C. An experiment in depth interviewing on the issue of internationalism vs. isolationism. *Publ. Opin. Quart.*, 1943, **7**, 267–279.
81. LORGE, I. The Thurstone attitude scales: I. Reliability and consistency of rejection and acceptance. *J. soc. Psychol.*, 1939, **10**, 187–198.
82. MATHEWS, C. O. The effect of printed response words upon children's answers to questions in two-

response types of tests. *J. educ. Psychol.*, 1927, **18**, 445–457.

83. May, M. A., & Hartshorne, H. First steps toward a scale for measuring attitudes. *J. educ. Psychol.* 1926, **17**, 145–162.

84. McNemar, Q. A critical examination of the University of Iowa studies of environmental influences upon the IQ. *Psychol. Bull.*, 1940, **37**, 63–92.

85. McNemar, Q. Sampling in psychological research. *Psychol. Bull.*, 1940, **37**, 331–365.

86. Merton, R. K. Fact and factitiousness in ethnic opinionnaires. *Amer. sociol. Rev.*, 1940, **5**, 13–28.

87. Miller, D. C. Morale of college-trained adults. *Amer. sociol. Rev.*, 1940, **5**, 880–889.

88. Miller, D. C. Personality factors in the morale of college-trained adults. *Sociometry*, 1940, **3**, 367–382.

89. Miller, D. C. Economic factors in the morale of college-trained adults. *Amer. J. Sociol.*, 1941, **47**, 139–156.

90. Miller, D. C. The measurement of national morale. *Amer. sociol. Rev.*, 1941, **6**, 487–498.

91. Miller, D. C. National morale of American college students in 1941. *Amer. sociol. Rev.*, 1942, **7**, 194–213.

92. Miller, D. C. Effect of the war declaration on the national morale of American college students. *Amer. sociol. Rev.*, 1942, **7**, 631–644.

93. Murphy, G., Murphy, L. B., & Newcomb, T. M. *Experimental social psychology.* New York: Harper, 1937.

94. Neely, T. E. *A study of error in the interview.* (Privately publ.), 1937. Pp. 150.

95. Newcomb, T. Labor unions as seen by their members: an attempt to measure attitudes. In G. W. Hartmann & T. Newcomb (Eds.), *Industrial conflict: a psychological interpretation.* New York: Cordon, 1939. Pp. 313–348.

96. Newcomb, T. M. *Personality and social change.* New York: Dryden Press, 1943.

97. Pace, C. R. A situations test to measure social-political-economic attitudes. *J. soc. Psychol.*, 1939, **10**, 331–344.

98. Proshansky, H. M. A projective method for the study of attitudes. *J. abnorm. soc. Psychol.*, 1943, **38**, 393–395.

99. Remmers, H. H., *et al.* Studies in attitudes. *Purdue Univ. Stud. higher Educ.*, 1934, **25**, 1–112.

100. Remmers, H. H., *et al.* Further studies in attitudes, Series II. *Purdue Univ. Stud. higher Educ.*, 1936, **31**, 1–298.

101. Remmers, H. H., *et al.* Further studies in attitudes, Series III. *Purdue Univ. Stud. higher Educ.*, 1938, **34**, 1–151.

102. Remmers, H. H., & Silance, E. B. Generalized attitude scales. *J. soc. Psychol.*, 1934, **5**, 298–312.

103. Reuss, C. F. Differences between persons responding and not responding to a mailed questionnaire. *Amer. sociol. Rev.*, 1943, **8**, 433–438.

104. Riker, B. L. A comparison of methods used in attitude research. *J. abnorm. soc. Psychol.*, 1944, **39**, 24–42.

105. Riker, B. L. Comparison of attitude scales—a correction, *J. abnorm. soc. Psychol.*, 1945, **40**, 102–103.

106. Roper, E. Wording of questions for the polls. *Publ. Opin. Quart.*, 1940, **4**, 129–130.

107. Roper, E. Checks to increase polling accuracy. *Publ. Opin. Quart.*, 1941, **5**, 87–90.

108. Rosander, A. C. An attitude scale based upon behavior situations. *J. soc. Psychol.*, 1937, **8**, 3–15.

109. Roslow, S., Wulfeck, H., & Corby, G. Consumer and opinion research: experimental studies on the form of the question. *J. appl. Psychol.*, 1940, **24**, 334–346.

110. Rugg, D. Experiments in wording questions: II. *Publ. Opin. Quart.*, 1941, **5**, 91–92.

111. Rugg, D., & Cantril, H. The wording of questions in public opinion polls. *J. abnorm. soc. Psychol.*, 1942, **37**, 469–495.

112. Rundquist, E. A., & Sletto, R. F. *Personality in the depression.* Minneapolis: Univ. Minn. Press, 1936.

113. Sanford, R. N., & Conrad, H. S. Some personality correlates of morale. *J. abnorm. soc. Psychol.*, 1943, **38**, 3–20.

114. Schanck, R. L. A study of a community and its groups and institutions conceived of as behaviors of individuals. *Psychol. Monogr.*, 1932, **43**, No. 2, 1–133.

115. Seashore, R. H., & Hevner, K. A time-saving device for the construction of attitude scales. *J. soc. Psychol.*, 1933, **4**, 366–372.

116. Shuttleworth, F. K. Sampling errors involved in incomplete returns to mail questionnaires. *J. appl. Psychol.*, 1941, **25**, 588–591.

117. Skott, H. E. Attitude research in the Department of Agriculture. *Publ. Opin. Quart.*, 1943, **7**, 280–292.

118. Stagner, R. Fascist attitudes: an exploratory study. *J. soc. Psychol.*, 1936, **7**, 309–319.

119. Stanton, F. Notes on the validity of mail questionnaire returns. *J. appl. Psychol.*, 1939, **23**, 95–104.

120. Stonborough, T. H. W. The continuous consumer panel: a new sampling device in consumer research. *Appl. Anthrop.*, 1942, **1**, No. 2, 37–41.

121. Suchman, E. A., & McCandless, B. Who answers questionnaires? *J. appl. Psychol.*, 1940, **24**, 758–769.

122. Thurstone, L. L. Attitudes can be measured. *Amer. J. Sociol.*, 1928, **33**, 529–554.

123. Thurstone, L. L., & Chave, E. J. *The measurement of attitude.* Chicago: Univ. Chicago Press, 1929.

124. Udow, A. B. The "interview-effect" in public opinion and market research surveys. *Arch. Psychol., N. Y.*, 1942, No. 277, 1–36.

125. Wang, K. A. Suggested criteria for writing attitude statements. *J. soc. Psychol.*, 1932, **3**, 367–373.

126. Watson, G. (Ed.) *Civilian morale.* Boston: Houghton Mifflin, 1942.

127. Watson, G. Morale during unemployment. In G. Watson (Ed.), *Civilian Morale.* Boston: Houghton Mifflin, 1942. Pp. 273–348.

128. Whisler, L. D., & Remmers, H. H. Liberalism, optimism, and group morale: a study of student attitudes. *J. soc. Psychol.*, 1938, **9**, 451–467.

129. Wilks, S. S. Representative sampling and poll reliability. *Publ. Opin. Quart.*, 1940, **4**, 261–269.

130. Wilks, S. S. Confidence limits and critical differences between percentages. *Publ. Opin. Quart.*, 1940, **4**, 332–338.

131. Woodward, J. L. Making government opinion research bear upon operation. *Amer. sociol. Rev.*, 1944, **9**, 670–677.

132. Young, K. (Ed.) *Social attitudes.* New York: Holt, 1931.

133. Zubin, J., & Gristle, M. An empirical scale for measuring militarism-pacificism. *Psychol. Rec.*, 1937, **1**, 27–32.

REPLY TO ELDER'S NOTE ON DUNLAP'S REMEDY FOR COLOR VISION

KNIGHT DUNLAP
University of California (Los Angeles)

In the note by Elder,* I found one item with which I agree completely, namely "The difficulties of conducting research from an academic base on a shifting war-time population." The difficulty was extreme, but we continued our research as a contribution to the armed services and the men therein.

As to our use of the American Optical Company's chart-test in spite of its unreliability, we used it because it was the test the men were required to pass for acceptance in certain of the services.

Concerning the Louisiana State University experiment, I did not reject the results because of unknown quantities of vitamin A, but because, as indicated in Elder's preliminary report (not included in the bibliography of the *Note*) the material was not pure vitamin A, but provitamin A.

We have long known that a "color-blind" person is unable to transform that provitamin into the vitamin, and we have had to abandon our research on so-called "color-blindness" because of the impossibility of obtaining vitamin A. Some pharmaceutical firms label their product Vitamin A, in large print, and in smaller print admit that it is a mixture of vitamin A, without indicating the proportions of the two factors. That the products of other firms, labeled Vitamin A, are mixtures of the same sort, is quite probable, even if it is not admitted.

As to Elder's suggestions for my future work: (a) was fulfilled in the work reported; (b) supervised administration was difficult for men in the services; (c) complete treatment was applied to the point at which some passed the service test; and (d) there was no possibility of the persons learning the tests, because they were not told whether their readings were right or wrong, or whether their judgments on the Nela test were right or wrong. As to Elder's statement that "on the basis of results presented on tests before and after treatment, he should have rejected 14 out of 14," I am not in agreement. As to the preceding statement: "If the summary statements given on each case may be interpreted literally then Dunlap himself rejects 12 of the 14 cases as questionable." I still disagree with Elder. Improvement was shown in almost all cases, and since the purpose of the treatment was to enable the individuals to pass the service tests, we could not go beyond that point. This reenforces my opening quotation from Elder's *Note*. The difficulty was indeed great.

---

* Elder, J. H. Note on Dunlap's remedy for color vision. *Psychol. Bull.*, 1946. 43, 77–79.

BOOK REVIEWS

Wertheimer, Max *Productive thinking*. New York: Harper & Bros., 1945. Pp. xiv+224.

America often reaps a harvest in providing haven for refugees. This time America—and the world—is richer for Max Wertheimer's studies carried out at the New School for Social Research. His book consists largely of concrete illustrations of "genuine, fine, clean, direct productive processes" given in great detail. In chapters I through III Wertheimer treats three relatively simple mathematical problems: finding the area of a parallelogram, proving the equality of the vertical angles formed by two intersecting straight lines, and Gauss' theorem. In chapter V he recounts the development of his own discoveries of the sum of the angles of a polygon, both plane and solid. Chapters VI and VII are psychological reconstructions of Galileo's discovery of the law of inertia and Einstein's thinking that led to the theory of relativity. In the concluding chapter Wertheimer summarizes the dynamic nature of the creative process, which he had slowly elaborated as he treated the various examples, as follows:

> When one grasps a problem situation, its structural features and requirements set up certain strains, stresses, and tensions in the thinker. What happens in real thinking is that these strains and stresses are followed up, yield vectors in the direction of improvement of the situation, and change it accordingly.

The problems thus far mentioned, with the exception of Einstein's, are relatively self-contained, closed systems. In chapter IV Wertheimer considers two concrete social situations in which the field is no longer limited to the internal forces in the objective situation, but includes concrete ego-tendencies and personal needs. In such cases, the solution arises out of transformations of the relation between the problem-situation and the ego. In each transformation, he claims the "structural features remain decisive."

On occasion, after posing a question suggested by the concrete problem under consideration, he answers in terms of logic and association theory. Then he demonstrates their inadequacy. Immediately he follows with an explanation based upon the structural characteristics of the situation and shows how the central core of the question now first becomes answered. Although the traditional theories at times are "adequate and useful," Wertheimer believes they are generally blind to the dynamics in thinking. They "show a dead picture stripped of all that is alive in them."

This book is offered to the psychologist, logician, and educator "as an invitation to reconsider basic issues." It is loaded with provocation. The educator is challenged to teach so that "genuine discovery in tasks of a structural nature plays an essential role." Emphasis on mechanical drill is "dangerous because it easily induces habits of sheer mechanized action, blindness, tendencies to perform slavishly instead of thinking, instead of facing a problem freely." The logician too is often taken to task for his structure-blindness.

There is a sparkle and warmth about this book that is captivating. Rarely does the psychological writer actually use the principles he extolls in the very construction of his book. Wertheimer does. For instance, he invites penetration, "Think it over yourself reader . . . Try to do justice to the points I am going to mention." The diagrams and illustrations themselves are used with remarkable effectiveness. The printer has placed them so expertly in the text that few formal references to them are required. They integrate smoothly into the flow of thought in the passage. Perhaps the quality which preserves the book's freshness from beginning to end is Wertheimer's attitude that he has highly satisfying and enjoyable insights to share with the reader. The book is gaily decked with such sentences as, "Try to do it before reading what I report in this chapter. You may enjoy what follows more."

Yet, when you have finished the book, you have no "clear" insight, you know that the dynamics of productive thinking are still far from being structurally "clean." No attempt is made to define concisely such terms as "inner structure." Wertheimer recognized and humbly admits this: "My terms should not give the impression that the problems are settled; they themselves are loaded with—I think—productive problems."

Scattered throughout the book are many suggestions for such research work. Wertheimer himself refers to a few odds and ends of systematic research he carried out. He does not give his methods or results explicitly, however, as he centers his full attention throughout the book upon the qualitative aspects of the productive process itself. It is curious that he never once refers to Duncker's *Zur Psychologie des productiven Denkens* (1935), which bears such close resemblance to his book.

The main value of Wertheimer's book will rest in its live challenge to psychologists to work in the field of creative thought, stressing the dynamic aspects. He wants contemporary psychologists to "look at the situation freely, open-mindedly, . . . to penetrate, to realize and trace out the inner" relationships of the thought process. The crucial test of the power of Wertheimer's invitation will be found in the number and quality of research workers who accept his challenge. Let us hope there will be many, for our world needs "clean" thinking. Wertheimer has done a remarkable job of blazing the trail; the way is no longer uncharted.

HAROLD GUETZKOW.

University of Michigan.

LECKY, PRESCOTT. *Self-consistency, a theory of personality.* New York: Island Press, 1945. Pp. 154.

This posthumous collection of essays, edited by John Taylor and supplied with an appreciative foreword of the author and his special psychological contributions by Gardner Murphy, develop the radical proposition that the individual "defines" for himself the nature of the totality which he is. New experiences are assimilated only when they are viewed as forward steps continuing and fulfilling oneself. This basic thesis rejects all the customary versions of the learning process, makes the organism its own determiner, and accepts but a single "general factor," viz., the personality itself.

Inventory data derived from Columbia College students indicate that confidence in social situations, optimism, a friendly attitude toward other members of the family, companionship with the opposite sex, efficient work dispositions, freedom from nervous symptoms, and a feeling of physical well-being are closely interrelated and tend to increase and decrease together. This is construed as fatal to the specificity doctrine. Instead, predictability of behavior is a function of stability which arises out of the basic need or general motive for consistent self-organization—the organism's sole purpose.

This position is supported by interpreting *learning* as always a resolution of conflict (reduction to unity), *pleasure* as difficult unification of a manifold (the barely successful synthesis), *emotion* as evidence of the effort required to achieve or preserve unity, and *personality* as an organization of values consistent with each other. Freud is critized for not recognizing that the patient's resistance is prerequisite to his integrity; *therapy must shift from the effort to make the patient consistent with society to making him consistent with himself.* Educational and allied defects are viewed as consequences of inhibiting "definitions" of our own natures from which we can be freed as soon as we see them as burdens and not as assets.

Lecky's clinical practice was apparently quite successful in using this clear-cut approach with its direct appeal to the client's standards of "belongingness." He asks psychologists to weigh the implications of their own "system-building" as evidence for his position. "The problems of science are not abstractions, but the human problems of scientists. If we did not strive for consistency in ourselves, what possibly could be the motive for seeking and demanding consistency in abstract scientific formulations?" (112).

These meaty papers are rich in both critical ideas and creative concepts. Lecky's abbreviated analysis resembles Allport's extended treatment at a number of points minus his trait theory and particularly the "eclectisicm" which is prima facie "inconsistent." Its positive features are an appealing simplicity, optimism about the possibilities of redirecting human energies, and ready availability for the tasks of the practitioner.

Nonetheless, grave skepticism is warranted at many points. Is the delinquent primarily such because he conceives of himself as a "bad boy"? Does one become honest (or a charming wife, or a keen lawyer, etc.) merely by affirming that such shall hereafter be part of one's nuclear self? Functional inadequacies here are purely peripheral; but what if the "central core" is diseased? The claim that one is or becomes largely what one "pictures" onself to be does not seem to square with the grim facts of frustration. Had Lecky lived to complete this theoretical structure, these questions might have received appropriate consideration. As it is, we must be grateful to him for his effective insistence that all individual behavior is a story with a plot wherein the evolving value-pattern of the actor is the most potent single psychological reality.

George W. Hartmann.

Teachers College, Columbia University.

Beck, S. J. *Rorschach's Test: 1. Basic processes.* New York: Grune & Stratton, 1944. Pp. xiii+223.

It is unquestionably true that Dr. Beck has once more demonstrated his competence and scholarship by producing this book. The question, however, is just how much is contributed by it—and by other like material—to American psychology and psychiatry. From where this reviewer sits, the answer to the question is disappointing.

Volume 1 of Beck's *Rorschach's Test* proposes to bring the presumptive reader up to date in Dr. Beck's cherished ambition to standardize and "demonstrate the processes used in evaluating Rorschach test responses." It fulfills its announced intention by painstakingly and deliberately detailing responses, rationalizing the use of the various symbols, explaining and cataloguing, arguing and justifying, through sixteen laborious chapters chaperoned by a brief bibliography, two appended tables, and ten disjointed schema of the blots. With commentary added to commentary, as one reads it, it takes on more and more the character of the Babylonian *Talmud*. To read such a book is obviously out of the question: to study it requires a fixity of purpose amounting almost to a dedication. Its real service will be as another standard reference work in a young but already needlessly over-complicated field.

The practicing psychiatrist or psychologist is coming to rely upon the Rorschach as he does upon almost no other tool. It is a good technique, perhaps the best of the projective methods, and he appreciates what it can do for him by way of evaluating the personality and, indeed, for certain services nowhere else available. He regards it practically, as an invaluable instrument for diagnosis, as an index and guide to therapeusis, as a clue to treatment progress. Hence the cultism of Rorschachers, their allegiances, their bickerings and jealousies, bore and distract him; so also do their endless disputes over—let us say—the criteria of distinction between one F and another. What the clinician wants from the Rorschach worker and researcher are ways of increasing the sensitivity of the test, enlarging its scope, and making it more informative. The present book, on the other hand, leaves the field where it was before it was written.

Perhaps the implied criticism of this review is unjustified in view of the fact that the volume under consideration really purports to do nothing more than reveal "basic processes." If this is so, however, then it must be taken as but another symptom of that compulsive overzealousness to substitute cold hieroglyphs for warm humans that so diverts the stream of clinical endeavor in the behavior sciences.

As for the volume itself, from the point of view of sheer book-making it is a satisfactory job of work. While the material could have been arranged in a manner more considerate of the reader and more sparing on his patience, it must be considered that merely setting up (much less proof reading) the text was a staggering accomplishment.

For those who are preoccupied with scoring problems in Rorschach work and who adhere to the Beck method as exposited already in his *Manual* and other published works, this volume is an amplification and extension of what they already know. But for those who are seeking help in their efforts to help others, this is merely another book bound in blue.

Baltimore, Maryland. ROBERT M. LINDNER.

WOLBERG, LEWIS R. *Hypnoanalysis*. New York: Grune & Stratton, 1945. Pp. xviii+342.

The hypnosis dealt with in this sober yet provocative work is by and large experimental hypnosis applied to the therapeutic situation, rather than the hypnosis of a purely suggestive treatment. The method includes slow induction of a deep trance, a supervening uninterrupted sleep for about 15 minutes, an hypnotic period comparable in activity to the waking state, during which one or more of the following procedures may be used: Free association to order, dreams to order, interpretation of his own dreams and of other behavior, age-level regression, experimental neurosis, automatic writing, crystal gazing, graphical representation of conflict material, play therapy, dramatic acting, giving of post-hypnotic suggestions, and finally a restful period of uninterrupted sleep. This hypnotic activity seeks not only to discover the traumatic episodes ("enucleation" of them) and the steps whereby false attitudes toward reality (hallucinations and delusions) have arisen from the patient's projection of his inner frustrations and needs ("empounded self-assertion"), but also to give practice in self-evaluation and even in self-assertion.

Because in hypnoanalysis transference with the stronger personality is indigenous, resistance can be understood and overcome. Since the unconscious strivings and attitudes are nearer to hand, the reality sense can be tested, and all in all, development of the ego can be actively fostered, it becomes an important adjunct to psychoanalysis, or as Dr. Wolberg declares, "Indeed, hypnoanalysis *is* psychoanalysis, performed in a controlled setting." Really, hypnoanalysis is a catalyst by means of which the psychoanalytic dynamisms precipitate out a cure in a short analysis.

Granted a skilled worker, the main limitations of hypnoanalysis are that not all patients can be hypnotized and not all hypnotizable persons have the ego strength to tolerate the transference. The fact that a patient may act out his inner impulses but fail to verbalize them constitutes a hazard (278).

In criticism of this excellent treatise the reviewer would call attention to two points. The first is that hypnosis, as described, depends on what the hypnotist makes of it and, no less, on how the subject reacts to it on the basis of his whole range of interpersonal attitudes (277). It is passive or active, enslaving or liberating, as the two participants consciously or unconsciously wish it—or it may be a stalemate.

The second point is that no new experimental proof is offered for the actuality of age-level regression. Negative findings, such as those reported in 1940 by the reviewer, are brushed aside with a reference to the consensus of opinion, with descriptions of cases, and with a long quotation from Erickson and Kubie describing their second stage of regression as so real that "the hypnotist and the hypnotic situation, as well as many other things, become anachronisms and nonexistent." Certainly, the existence of such a self-consuming stage must be verified under the best of experimental controls.

PAUL C. YOUNG.

Louisiana State University.

TREDGOLD, A. F. *Manual of psychological medicine for practitioners and students* (2nd Ed.). Baltimore: Williams & Wilkins, 1945. Pp. xi +308.

Psychological medicine is concerned with disease of the mind as apart from body. Mind is manifested as conation, affect, and cognition. The first two provide the activating force of conduct. Conduct abnormality is classified as disorder, decay, and defect, each having clinical varieties. Disorder and decay may be primary (functional) or secondary (organic). Defect may be primary (germinal) or secondary (environmental). Disorders arise when an inherited disposition is subjected to environmental stress; in decay inherent defective durability is the principal causal factor. The major portion of the book is devoted to the incidence, symptomatology, etiology, prognosis, and treatment of the clinical varieties. In a general discussion of treatment, psychotherapy is described as explanation, persuasion, and suggestion. There is an adequate index, a 52-item bibliography, but there are no illustrations.

The principal weakness of the book is its outmoded dualism and its consequent perpetuation of a misconception of the functional psychoses as presumably involving no physiological function. The discussion of medical treatments is generally up-to-date but there is no mention of refrigerants or of diet control in epilepsy. Nearly 65 per cent of the bibliographical references antedate 1930.

The first edition (1943), written to fill the wartime need of British medical officers, was exhausted in nine months, hence this edition, which differs solely in a small number of additions, has appeared. Its brief descriptions of the rarer psychoses will make it useful as a reference. However, other adequate sources free from Tredgold's dualistic approach are already available.

Illinois State Normal University. STANLEY S. MARZOLF.

KVARACEUS, W. C. *Juvenile delinquency and the school.* Yonkers: World Book, 1945. Pp. x+337.

Whether or not juvenile delinquency is a seriously increasing problem cannot be clearly answered with available data; but that juvenile delinquency has been for many years a serious social problem cannot be denied. In the final analysis it is only study of delinquents as individuals which can lead to understanding and perhaps ultimate solution of the problems. The volume being reviewed is the latest in a long series of studies starting with Healy's *The Individual Delinquent* in 1915 which present basic data on the psychological, social, and other characteristics of delinquents determined from studies of individuals.

In 1937 the Children's Bureau was started under the general direction of the Passaic, N. J. Board of Education. Its interest is in truancy and delinquency with the purpose "to eliminate the causes of these offenses, to prevent their occurrence, and to make desirable adjustments for children who have offended." Dr. Kvaraceus, who has been Director of the Bureau, has analyzed the data on the first 761 cases and presents reasonably exhaustive findings in such areas as early development, family, economic status, the community, and the school. In general his findings are in agreement with earlier studies but the group is somewhat unusual in that it contains a relatively high proportion of pre-delinquents.

The recurrent emphasis in this book concerns the place of the school in the problem of delinquency. The author has no hesitancy in pointing out that frustrating experiences in schools have etiologic significance in delinquency. But he argues well that community programs of dealing with the problem have much to gain if the schools will accept leadership. Certainly the work here described, while not perfect, gives sufficient evidence of success to warrant close study, and imitation, by other communities.

While this work is not particularly oriented by psychologists it will be of value to all those dealing with the delinquent. Its greatest value is as a stimulus to community thinking on a community problem.

C. M. LOUTTIT.

Ohio State University.

BRANDT, HERMAN F. *The psychology of seeing.* New York: The Philosophical Library, 1945. Pp. xvi+240.

This is a difficult book to assess adequately as a unit, because it is essentially a collection of experimental articles loosely tied together with somewhat casual generalizations. When considered separately, many of the individual studies have considerable merit. All of them are based upon a single piece of apparatus, a special camera, invented by the author, which photographs simultaneously the horizontal and the vertical movements of the eyes separately on the same film at points a fixed distance apart. To discover how this is accomplished one must turn to an earlier article listed in the bibliography but not referred to in the text. Illustrations are given, however, of the final projection of the records so that a picture of the excursions of the eyes over the test object is obtained.

There follows a discussion of *Basic Eye Movement Tendencies* with no reference in the text to the rather considerable literature on eye movement. Appending a bibliography hardly compensates for this lack. The reason may be found, perhaps, in the preparation of the original articles for non-technical and semi-popular readers.

Ocular photography is applied, then, to problems of advertising which encompass attention, interest, and preferences; to problems of learning, of art and of perception. Finally, a program of projected studies is laid down, ranging from the measurement of intellectual abilities and aptitudes through determination of the effects of fatigue to detection of guilt or innocence of criminals. The applicability of ocular photography is boundless.

Presentation of the various studies tends to follow a pattern, viz., problem, apparatus, procedure and results. Generalizations and interpretations are made frequently but do not always follow from the data. The attempt to consolidate the various units into a general psychological treatise seems forced and casual. The author's enthusiasm is for the apparatus and its applicability, and this leads to unnecessary repetitions of procedural descriptions and exclamations of satisfaction with the results. Four separate illustrations of the camera are shown, all from so nearly the same angle that they are essentially identical. Such an external view reveals none of the essential features of the device.

Perhaps the reviewer has approached this book from too professional an angle. A suggestion that this is the case is given in the glossary which defines such terms as area, art, color, distraction, fatigue, learning, retina, sensation, etc. Criticism, however, is disarmed by the author's statement in the preface, "The readers may at times sense a repetition of methods of procedures employed in studies assembled and occasional sweeping statements. The common pattern emerged as a result of an attempt to combine studies published elsewhere, and if over-enthusiasm is manifest it is largely because the writer projected his imagination beyond known facts for the purpose of citing potential fields of investigation."

Nevertheless, the generality of the title is misleading. While the problems reported are primarily psychological, they do not, all together, constitute "Psychology" nor all of "seeing." A more definitive title would give a better impression.

FORREST LEE DIMMICK.

Hobart College,
Geneva, New York.

Vol. 43, No. 5 September, 1946

Psychological Bulletin

478z88

THE VALIDITY OF PERSONALITY QUESTIONNAIRES

ALBERT ELLIS
2505 University Avenue, New York 63, N. Y.

While the reliabilities of personality questionnaires have usually been notoriously high, their validities have remained more questionable. Indeed, some of the most widely known and used paper and pencil personality tests have been cavalierly marketed without any serious attempts on the part of their authors to validate them objectively. Half-hearted "validations" in terms of the internal consistency of the tests have, to be sure, usually been offered; but no real endeavors have been made to show that, when used according to their standard directions, these instruments will actually do the clinical jobs they are supposed to do: meaning, that they will adequately differentiate neurotics from non-neurotics, introverts from extroverts, dominant from submissive persons, and so on.

In consequence, a cursory survey of the literature on personality questionnaires will show that for every validity study that has been made there are perhaps three or four reliability experiments; and that of the relatively few validations that have been reported, the majority are not objective, or clinical, validations at all, but are essentially little more than statistical checks and balances on the particular method of test construction that has been employed. Naturally, these internal consistency "validations" rarely fail.

In order that a clearer picture of the validity of personality questionnaires may be obtained by both the proponents and opponents of this technique for diagnosing human behavior, we propose to summarize the findings of those studies which have reported clinical validations—studies, that is to say, where the test or tests under investigation have been tried on diagnosed groups of one type or other, to see whether they distinguish the individuals in these groups as they are supposed to do.

THE GENERAL CASE AGAINST THE VALIDITY OF PERSONALITY QUESTIONNAIRES

Ever since Woodworth brought forth his Personal Data Sheet in 1917, psychologists have never hesitated to point out defects in personality schedules and to question their validity. Among the points that have been raised against paper and pencil personality tests have been these:

1. They do not bring out configurational meanings, like the case study method of diagnosis presumably does, and hence can give no whole, or organismic, picture of human behavior (Alexander (**4**); Kirkpatrick (**157**); Layman (**175**)).
2. They may sometimes give an adequate picture of a group of persons, but rarely can be of use in individual diagnosis (Frank (**86**); Moore and Steele (**218**); Spencer (**291**); Terman (**307**)).
3. Personality cannot be described in terms of single traits, such as neuroticism. Questionnaires which purport to estimate such a trait are really measuring a composite of several different ones (Guilford (**104**); Mosier (**220**)).
4. The questionnaire is sometimes unreliable, and when this is the case it cannot have a high validity either (Bigham (**30**); Lentz (**176**)).
5. Some questionnaires (like the Bernreuter Personality Inventory) which purport to measure several distinct traits are, at best, measuring the same one under two or more different names (Lorge (**181**); Perry (**235, 236**)).
6. Cultural factors, which cause persons raised in different communities to make different responses to the questionnaires, may interfere with their validity (Shen (**277**)).
7. Questions on personality schedules are interpreted in different ways by different individuals, and hence make test results suspect (Eisenberg (**71**)).
8. There is a general over-estimation, or self-halo, effect when normal persons take personality questionnaires (Allport (**9**); Laslett (**174**)).
9. Most subjects can easily falsify their answers to paper and pencil tests of personality, and frequently choose to do so (Allport (**9**)).
10. Questionnaires are administered under all sorts of conditions, and the motivations of their respondents may vary in consequence. True clinical rapport is not possible under such circumstances (Eysenck (**72**)); Frank (**85**); Maller (**200**); Vernon (**329**)).
11. The manner in which responses must be made in questionnaires—for example, on a YES ? NO or on a rating-scale basis—may influence the validity of schedules (Lorge (**183**)).
12. Lack of internal consistency may invalidate a questionnaire, but presence of internal consistency will not necessarily validate it (Lorge (**182**)).
13. The manner in which the directions of a test are given may significantly affect the results made on it (McQuitty (**194**); Van Wagenen (**328**)).
14. The particular way in which questions are stated, and their vocabulary range, may cause many misunderstandings by the respondents

and thus adversely affect test validity (Hampton (**111**); Papurt (**233**); Rundquist (**265**)).

15. Testing is an artificial procedure which has little to do with real-life situations, and its results are consequently not very valid (Krait (**161**); Watson (**337**)).

16. Some personality questionnaires are "validated" against other questionnaires from which, in turn, their items were largely taken, thus rendering their "validations" spurious (Kuznets (**163**)).

17. Even when the respondent does his best to answer questions truthfully, he may lack insight into his true behavior or may unconsciously be quite a different person than the picture he draws of himself on the test (Alexander (**4**); Laslett (**174**)).

18. Most questionnaires consist of a pot-pourri of items of much or little or no significance, so that no one quite knows what their total scores really mean (Watson (**337**)).

19. Chance plays a large factor in the way many questionnaires are checked and scored (Wile (**345**)).

20. While low (or *neurotic*) scores on personality schedules may mean something, high (or *non-neurotic*) scores may mean nothing at all, and may, in fact, indicate serious maladjustment (Johnson (**149**); Landis & Katz (**172**)).

21. Armchair, rather than empirical, construction and evaluation of test items is frequently the case in personality questionnaires (Landis, Zubin & Katz (**173**)).

22. Uncritical use of statistical procedures with many personality tests add a spurious reality to data which were none too accurate in the first place (Landis (**169**)).

23. Questionnaires of the YES NO type often force inaccurate judgments on their respondents, since there are no intermediate answers they allow (Landis (**169**)).

24. Many personality tests which claim to measure the same traits—*e.g.*, introversion-extroversion—have very low intercorrelations with each other (Flemming (**83**); Guthrie (**110**); Stagner (**293**)).

25. There are no statistical short-cuts to the understanding of human nature, such as the ones many test makers try to arrive at through involved factorial analyses (Burks (**39**); Terman (**307**)).

26. When used for occupational purposes, personality questionnaires give their respondents incentives to over-rate themselves, and are consequently usually invalid (Traxler (**320**); Vernon (**329**)).

Because of objections like the foregoing ones, many psychologists have been led to state unequivocally that paper and pencil personality schedules have, at present, a low, or at least a disappointing, degree of validity. For example: Gilliland (**93**); Maller (**196, 200**); Roback (**251**); Rosenzweig (**259**); Thorpe (**311**); Traxler (**320**); Vernon (**329**); and Wiley and Trimble (**346**). Confirming these impressions of the dubious worth of personality tests, a recent study by Kornhauser (160) showed that the majority of leading psychologists hold these instruments in low repute. Kornhauser received answers from 67 well-known psychologists to the question: "In the field of personality testing, how satisfactory or helpful for present practical use do you consider Personality Inventories and questionnaires (such as those of Bernreuter, Bell, Humm-Wadsworth, etc.)?" Only 1.5% of his respondents deemed these instruments *highly satisfactory;* 13.5% considered them *moderately satisfactory;* 36%, *doubtfully satisfactory;* 33%, *rather unsatisfactory;* and 16% *highly unsatisfactory.*

THE CASE FOR THE DEFENCE OF THE PERSONALITY QUESTIONNAIRE

In spite of the many assaults that have been made against it, the paper and pencil personality test has got along splendidly as far as usage is concerned. For there can be little doubt that Americans have, to date, taken more of the Woodworth-Thurstone-Bernreuter type of questionnaires than all other kinds of personality tests combined. Some of the reasons for this widespread use may be found in these points which are often made in its favor:

1. Personality questionnaires are standardized instruments that can be handled in an objective, statistical fashion. (Hunt, Wittson, and Harris (142)).
2. They are usually very easy to administer and score (Maller (200)).
3. They almost always have some degree of validity, since even when their respondents state that they do things which they really don't perform (or *vice versa*), the fact that they *state* that they act in such a manner is an important one for the understanding of their personalities (Garrett (91); Meehl (214)).
4. In the hands of an experienced clinician, personality tests may give valuable information about their respondents. Even though total scores may be somewhat meaningless, examination of individual responses may provide valid clinical material (Gorham and Brotemarkle (97); Hunt, Wittson and Harris (142)).
5. Statistical analysis shows that the traits posited by questionnaires have a real existence and are not the result of chance factors (Bernreuter (28)).
6. Investigation shows that normal and abnormal subjects do not differ with respect to the interpretations they give to questions in personality schedules (Benton (21)).
7. Personality testing is still very young, and as it becomes more refined may yet throw much light on our understanding of human nature (Richmond (248)).

8. Psychometric analysis of personality may yet replace the clinical approach as commonly practiced for diagnostic purposes today (Thomas (309)).

9. It does not matter if respondents answer untruthfully on personality questionnaires, since allowances are made for this in the standardization or scoring of the tests (Symonds (301)).

10. The questionnaire may be employed as a formalized interview, and may thereby given valuable personality information (Durea and Bilsky (70); Landis (169)).

11. Traditional methods of validating questionnaires by outside criteria are themselves faulty and invalid; hence, validation by internal consistency alone is perfectly sound (Link (178); May (210); Root (255); Stagner (296)).

It can be seen, then, that there is something to be said on the side of paper and pencil personality tests. Indeed, in the last few years there has been something of an upsurge of interest in these questionnaires, largely resulting from the application of factor analysis and other new statistical procedures to their construction and standardization. Work along these lines has been reported by such authors as Darley & McNamara (66); Guilford & Martin (100, 102, 104, 105, 106, 107, 108); Flanagan (80, 81); Gibb (92); Harris and Dobelstein (116); Lorge (181, 182, 183, 184); Martin (206); Mosier (221); Thurstone (312); Wolfle (352); and Zubin (357).

PREVEWS SURVEYS AND REVIEWS OF THE LITERATURE

Since this article reviews only the literature on the validity of personality questionnaires, it might be well to mention here some of the surveys and bibliographic reviews which cover the general field of personality testing. These include studies by Allport (6); Allport and Vernon (10); Buros (40, 41, 42, 43); Cattell (50); Conklin (54, 55); Darley and Anderson (64); Freeman (88); Guilford and Braley (103); Guilford (101); Hildreth (126); Horsch and Davis (129); Koos (159); Maller (196, 197, 199, 200); May, Hartshorne, and Welty (211, 212, 213); Murphy (223); Roback (256); Snyder (288); Symonds (301, 303); Traxler (318, 320, 321, 322); Watson (335, 336, 337, 338, 339); and the National Educational Association (358, 359).

The closest review to this one in the literature is that of Super (298, 299) in the *Psychological Bulletin* of 1942. He covers the validity studies on the Bernreuter Personality Inventory up to that date, and concludes: "The Bernreuter scores tend to distinguish neurotics and various types of psychotics from normal individuals, although not perfectly; abnormals in the mass are not differentiated, because of the canceling out of extreme tendencies. Unfavorable scores are indicative of maladjustment, but 'favorable' scores do not necessarily indicate good adjustment . . ." (**298**, p. 115).

CRITERIA OF VALIDITY EMPLOYED IN THIS ARTICLE

In general, there are two methods of validating personality questionnaires: a *subjective* and an *objective* one. The *subjective* method consists of checking the test against itself: that is, seeing whether respondents answer its questions in a manner that shows it to be internally consistent. What validation by internal consistency boils down to, in frequent practice, is the test maker's rather arbitrarily deciding that certain answers to the questions he selects for his test will be scored as *bad* or, let us say, *neurotic* ones. He then gives the test to a mixed group of respondents and diagnoses as *neurotics* those individuals who consistently give *bad* answers to it. Thus, he may select as *neurotic* those who score in the lowest decile or quartile of the test and as *non-neurotics* those who score in the highest decile or quartile. Subsequently, he checks back, often with involved statistical procedures, to make sure that the individual items in his schedule actually differentiate his test-diagnosed *neurotics* from *non-neurotics*; and he keeps only those items which do this fairly well. Then, perhaps, he gives his revised test all over again and demonstrates that his *neurotics* still get high *neurotic* scores: thereby "validating" the test.

The *objective* method of personality test validation consists of checking a schedule, preferably item by item, against an outside clinical criterion. Thus, a questionnaire may be given to a group of normal individuals and to another group of subjects who have been diagnosed by competent outside observers as *neurotic*, or *maladjusted*, or *psychotic*, or *delinquent*, or *introverted*. If, then, the clinically diagnosed group makes significantly higher *neurotic* scores than does the normal group, the test under consideration is said to have been validated; and if there are no significant differences between the two groups, the validity of the test remains, to say the least, in doubt.

The disadvantages of both these methods of validating personality schedules are fairly obvious. The *subjective* validations may be strongly questioned as to their constituting validity, in any strict sense of the term, at all. A test of neuroticism, to be truly valid, must do what it is supposed to do: that is, clinically differentiate neurotics from non-neurotics. Even if the test-maker carefully selects his items by preliminary investigation and experimentation; even if he retests his validation group (instead of simply carelessly re-scoring their papers); and even if he employs for final validation purposes a different group than the one originally employed for item-selection, he still cannot say whether his test is *clinically* valid. Internal consistency of a questionnaire demonstrates, at best, that it is a reliable test of *something;* but that *something* may still have little or no relation to the clinical diagnosis for which the test has presumably been designed.

Moreover, the statistical assumptions underlying the use of the technique of internal consistency for test validation are themselves subject to debate. Thus, Jackson and Ferguson (**146**) have questioned whether this technique even gives a true test reliability; and Rundquist and Sletto (**266**) have pointed out its dangers as a validation procedure, and have emphasized the difficulty of obtaining suitable degrees of separation in the original criterion groups selected for item analyses.

On the other hand, *objective* or clinical validations of personality questionnaires also leave much to be desired in most cases. For these validations depend, in one way or another, on diagnosis by ratings, whether it be the ratings of psychiatrists, clinical psychologists, friends, associates, teachers, supervisors, or other tests of personality. Consequently, questionnaires may be *objectively* validated against psychiatric diagnoses which are highly questionable, against friends' or teachers' ratings which widely differ among themselves, or against other personality testing instruments which are just as suspect as that one under investigation.

Objectively validating personality schedules against institutionalized cases—*e.g.*, psychotics or delinquents—is also far from satisfactory. For one thing, these groups are frequently subjected to answering questions, and their members may therefore become more inured to or sophisticated towards them than are normal groups. Secondly, institutionalized persons have already been punished, as it were, for their personality defects and have possibly less to hide than non-institutionalized individuals who may fear some form of penalization if they admit to possessing certain traits. (The reverse may also hold true in some special instances.)

It must also be pointed out that, to date, sampling and statistical procedures used by those employing *objective* validation methods have frequently left much to be desired in the way of scientific accuracy. Criterion groups have not been adequately selected or sub-divided. Rather meaningless correlation coefficients have been given, while more meaningful coefficients of dependability or "forecasting efficiency" have been suspiciously ignored. Significant differences in *group* means have been inferentially given a spurious connection with *individual* diagnosis that they in no wise truly possess. Thus, test-makers (and test-sellers) have pointed with pride to their instruments distinguishing between normal and neurotic individuals, when they really meant between *groups* of normals and neurotics—quite a different thing!

Admitting, therefore, the questionable aspects of *objective* methods of validating personality questionnaires, we shall nevertheless review in this paper the available studies using outside criteria of validity. We shall not include *subjective* validity studies because: (1) it is questionable, as we pointed out before, whether they are true validity experiments at all; and (2) practically 100% of those reported in the literature give positive, or "valid," results—thus making their enumeration pointless.

Considering the *objective* validity investigations, we shall review those which fall under the following six headings:

1. *Behavior Problem Diagnosis.*—Experiments where a personality questionnaire was given to normal children on the one hand and diagnosed behavior problem children on the other hand.
2. *Delinquency Diagnosis.*—Experiments where a test was given to normal children and to institutionalized (or otherwise accredited) delinquents.
3. *Psychiatric or Psychological Diagnosis.*—Experiments where a test was given to normal groups and to groups of individuals who were diagnosed by psychiatrists, clinical psychologists, or trained supervisors to be definitely maladjusted, neurotic, or psychotic.
4. *Rating diagnosis.*—Experiments where a test was given to subjects who were carefully rated by their friends, associates, teachers, or supervisors on the trait or traits whose diagnosis was at issue.
5. *Test Intercorrelations.*—Experiments where individuals were given two or more personality questionnaires presumably measuring the same trait or traits.
6. *Over-rating or Lying Validations.*—Experiments which were specifically designed to see whether respondents over-rated themselves or lied in their answers to personality questionnaires.

Only questionnaires of the Woodworth, Thurstone, and Bernreuter type will be considered in this review. Experiments involving tests like the Downey Will-Temperament or the Pressey X-O series will not be included, since they are not questionnaires of the usual type. Experiments using the Minnesota Multi-Phasic test will be considered in a separate section since, as it is ordinarily administered, this inventory is given individually, and not as a group test, as are other questionnaires.

Nearly all the scales employed in the following experiments are tests of *neuroticism* (*social adjustment*, *maladjustment*, *emotional stability*, *emotional affectivity*) or *introversion*, since these two traits have been found empirically to be almost identical as far as personality questionnaires are concerned, and since the validity of tests purporting to measure them is still very much in doubt. Investigations involving such scales as *health adjustment, dominance-submission, occupational adjustment*, and so on are omitted from the review because their validity has not usually been particularly questioned. Tests of interests and attitudes, opinion, moral values, character, and the like are also omitted here, since they do not generally come under the heading of paper and pencil personality questionnaires.

One of the difficulties of evaluating validity studies in this field is that of appraising coefficients of correlation reported in some of the studies. Chi-square, *t*, or *F* scores may easily be referred to probability tables for an estimation of their significance. But "significant" correlations usually merely mean correlations that significantly differ from *zero*, and are of limited interpretative value.

Strictly speaking, many correlations reported on personality-test validity experiments should not be shown in terms of *r* at all, but in terms of what Bingham (**30**) calls *D* (the coefficient of dependability) or what Garrett (**89**) calls *E* (the coefficient of forecasting efficiency). *D* or *E* is a measure of a test's actual efficiency in predicting criterion scores, and is always considerably lower than *r*. Thus, when *r* is .99, *D* is .86; when *r* is .90, *D* is .56; and when *r* is .70, D is only .29.

Since holding personality test evaluations to terms of *D* or *E* rather than *r* would probably be unfair at the present stage of their development, we shall, in this review, usually evaluate the reported coefficients of correlation in terms of the conventional estimations given them in the consideration of psychological and educational tests. Thus, we shall say that *r*'s of from 0 to .19 indicate *negative* validity; from .20 to .39, *mainly negative* validity; from .40 to .69, *questionably positive* validity; from .70 to .79, *mainly positive* validity; and from .80 to 1.00, *positive* validity.

Keeping the foregoing considerations and limitations in mind, we shall now review the literature on the *objective* validity of personality questionnaires.

VALIDATION BY BEHAVIOR PROBLEM DIAGNOSIS

Relatively few studies attempting to validate personality questionnaires by behavior problem diagnosis have been reported up to the present time. They include the following ones:

1. Daniel gave the Woodworth-Mathews test (**44, 87, 128, 209, 353**) to Negro behavior problem boys and to a group of normal Negro controls. He found a reliable difference in the scores of the two groups (**62**, p. 26). *Result:* positive.

2. Daniel gave the Maller Character Sketches (**115, 198**) to the same two groups and again found a reliable difference in the group scores (**62**, p.33). *Result:* positive.

3. Mathews gave the Bernreuter Personality Inventory (**24, 25, 26, 27, 29, 56, 74, 111, 152, 278**) to 93 boys in a six-year secondary school, and reported that five of its six scales failed to differentiate between good and poor behavior groupings of the students. The only scale that did show significant differences between the groups was the *sociality* one (**208**, p. 47). *Result:* mainly negative.

4. Riggs and Joyal gave the Loofbourow-Keys Personal Index (**153, 154, 180**) to 186 junior high school boys divided into best and worst groups, in terms of problem behavior. There was a coefficient of correlation of .48 between the scores of the boys and the final ratings of their advisers. "This is not a very high degree of positive relationship. . . . Perhaps not over one boy in three thus indicated is likely to become a problem behavior boy" (**349**, pp. 197–201) *Result:* questionably positive.

5. Speer gave the Bernreuter Inventory to 54 behavior problem children and 185 normal school children, and found that "there is no significant difference between the scores of the problem and non-problem groups on any of the four scales. . . . It would appear that the Bernreuter Personality Inventory is of no advantage in the prediction of problem children" (**289**, p. 69). *Result:* negative.

6. Symonds and Jackson gave the Symonds Adjustment Questionnaire (**302, 304, 306**) to high school pupils and found a correlation of .38 between their scores on this test and teachers' ratings of them as problem children. They concluded that "these instruments possess only limited diagnostic value" (**305**, pp. 95, 105). *Result:* mainly negative.

7. Thomson gave the Link P.Q. Test (**178, 179, 260, 261**) to 146 boys and 113 girls at a Mooseheart, Illinois high school and reported: "An apparent tendency for low PQ's to be associated with problem boys does not prove to be statistically reliable. No evidence was available to indicate a relationship between PQ and behavior problems among girls" (**310**, p. 288). *Result:* negative.

8. Yourman gave the Thurstone Personality Schedule (**119, 164, 246, 294, 313, 314, 348**) to children identified by their teachers as problems and found significant differences between their responses and those of normal children on only nine of its questions (**355**, p. 337). *Result:* mainly negative.

9. The present author, in an unpublished study, administered a questionnaire consisting of the most frequently used questions in the Woodworth, Maller, Symonds, Loofbourow-Keys and other schedules to 221 normal and 40 behavior-problem junior high school boys and found that it did not significantly distinguish between the two groups. *Result:* negative.

In sum: Of nine attempts to validate personality questionnaires with groups of behavior problem children, two have given positive, one questionably positive, and six negative results.

VALIDATION BY DIAGNOSIS OF DELINQUENCY

An examination of the literature produces the following instances in which experimental attempts have been made to validate personality questionnaires by comparing results obtained on them by groups of normal and delinquent children:

1. Ascher and Haven, using the Woodworth-Mathews Test on 594 public school boys and 249 inmates of the Kentucky House of Reform found that total

scores "revealed no significant differences between the two groups. . . . There was a significant difference in the responses of the correctional and public school boys to only six of the (93) questions" (**12**, p. 106). *Result:* negative.

2. Babcock used the Sweet Personal Attitudes Test (**300**) on normal and delinquent groups of boys in Rochester and New York stated that "no significant differences between the mean of the institution and the public school boys were found in any of the seven categories of this test, either among New York or Rochester groups" (**13**, p. 46). *Result:* negative.

3. Babcock employed the Rogers Personality Test (**254**) on these same groups and found that again there were no significant differences between the normal and delinquent boys (**13**, p. 50). *Result:* negative.

4. Bartlett and Haas employed the Bell Adjustment Inventory (**17, 18, 268, 326**) with 119 delinquents and 148 high school and grade school boys and found that although no significant differences existed between the two groups on its general social adjustment scale, "marked differences appeared in the degree of emotional adjustment" (**15**, p. 654). *Result:* mainly positive.

5. Bartlett and Haas used the Woodworth-Mathews test on the same groups and reported that "the delinquents made an average of 17.76 points, as compared with the non-delinquent boys' average of 10.21 points, indicating greater emotional instability in the delinquent group" (**15**, p. 654). *Result:* positive.

6. Boynton and Walsworth examined 50 normal girls and 47 inmates of the Tennessee State Reform School, and found a critical ratio of 2.33 between the differences of the scores of these groups on the Maller personality test (**32**, p. 86). *Result:* questionably positive.

7. Boynton and Walsworth used the B.P.C. Personal Inventory with the same groups and found a critical ratio of 2.12 between them (**32**, p. 86). *Result:* questionably positive.

8. Boynton and Walsworth reported the Bernreuter scales as achieving ratios ranging from .82 to 1.91 between the same delinquent and normal groups (**32**, p. 86). *Result:* negative.

9. Boynton and Walsworth found the four Rogers scales attaining critical ratios ranging from .34 to 4.56 for the same groups (**32**, p. 86). *Result:* questionably positive.

10. Bridges gave the Woodworth-Mathews questionnaire to 33 delinquent girls and reported that "these delinquent girls have as a group nearly twice as many of the abnormalities investigated [as have Mathews' standardization group]. This difference is, however, probably not as great as it seems, since the average age of the delinquent group is greater than that of the Mathews cases, and she had shown that for unselected girls the number of symptomatic responses increases with age" (**33**, p. 195). *Result:* questionably positive.

11. Bridges administered the Woodworth P.D. Sheet to the same groups and found that "the delinquent girls do not differ so greatly from the women students. Fifty per cent of the women students and 60 per cent of the delinquent girls have scores about 20, which in the dase of men is supposed to indicate marked instability" (**33**, p. 203). *Result:* mainly negative.

12. Bridges and Bridges gave the Woodworth-Mathews Test to 98 delinquent boys at the Boys' Training Farm School in Quebec, and found that "their median was 21 symptomatic responses as compared with nine for the unselected school boys studied by Mathews" (**34**, p. 455). They concluded that this was a significant difference. *Result:* positive.

13. Capwell gave the Washburne Social Adjustment Inventory (**332**, **333**) to delinquent and non-delinquent adolescent girls and reported that "54% of the delinquents reached or exceeded the 75th percentile of the non-delinquents' scores on the first examination and 51% did so on the second examination" (**46**, pp. 224–5; **47**, pp. 291–2). *Result:* mainly positive.

14. Casselberry gave the Neyman-Kohlstedt test (**236**, **294**) to 329 delinquent and 53 non-delinquent boys, finding that "delinquent boys are slightly more introverted than non-delinquents according to . . . this test" (**48**, p. 6). So little use was the test in discriminating delinquents from non-delinquents that he eliminated it from a diagnostic battery he was devising. *Result:* mainly negative.

15. Casselberry gave the same groups the Woodworth P.D. sheet and found that while "as a group the delinquents appear to be less stable emotionally according to this test," it did not have a high degree of discrimination. Again, he had to reject it from his diagnostic battery (**48**, p. 8). *Result:* questionably positive.

16. Casselberry used the Bernreuter with the same groups, and reported that "this blank failed to show a significant difference between the delinquent and the non-delinquent groups in any of the scores obtainable from it" (**48**, p. 17). *Result:* negative.

17. Cavan found that the White House Scale (a questionnaire consisting of questions from the Cady, Mathews, and Slawson revisions of the Woodworth) "differentiates clearly" between groups of adolescent delinquents and non-delinquents (**51**, p. 26). *Result:* positive.

18. Courthial gave the Woodworth-Mathews test to 147 delinquent girls in two reform schools and compared their scores to those of normal girls. His comment was that "the significance of the difference is more than twice what it should be in order to be perfectly reliable. Delinquent girls show . . . a tendency towards emotional instability and neurotic responses, such as are elicited by the test, to a much greater degree than well-adjusted girls" (**59**, p. 55). *Result:* positive.

19. Cushing and Ruch administered the Woodworth-Cady test to 127 girls in the Iowa State Training School and to a control group of 50 normal girls. They gave a critical ratio for the difference between the mean of the two groups of 8.8, which is highly significant (**61**, p. 4). *Result:* positive.

20. Daniel employed the Woodworth-Mathews sheet with normal and delinquent boys and found "a reliable difference . . . between the mean of the symptomatic responses . . . and the mean of . . . the delinquent boys," but also found that "an analysis of the responses to the specific items revealed that as compared with the non-problem group a larger proportion of the delinquents gave unfavorable responses in 58% of the items. . . . Statistically reliable differences between the percentages of unfavorable responses of the non-problem and delinquent boys were found in 13 items" (**62**, p. 26). *Result:* mainly positive.

21. Daniel found that, for the same groups, the Maller Character Sketches gave more reliable differences than did the Woodworth-Mathews test results previously noted (**62**, p. 33). *Result:* positive.

22. Harris tried the Maller Inventory on delinquent boys in a midwest state institution and reported that "whatever traits of personality adjustment

the Maller Inventory may measure, it does not appear to measure those traits which are related to successful behavior adjustment in the state school of this study" (**115**, p. 555). *Result:* negative.

23. Horsch and Davis found that the Bernreuter emotional stability (B1-N) scale differentiated reliably between normal high school students and delinquents in the Colorado State Industrial School. It did not, however, reliably distinguish State Reformatory inmates from college students of equivalent ages (**130**, p. 62). *Result:* questionably positive.

24. Houtchens used the Humm-Wadsworth Temperament Test (**133**, **134**, **135**, **136**, **137**, **138**, **139**, **162**) with delinquent and non-delinquent adolescents and concluded that "differences in delinquent and non-delinquent adolescents were found to be significant in all components except the epileptoid" (**132**, p. 61). *Result:* positive.

25. Landis found that the Thurstone Test "fails to give diagnostic results when revised for use with juvenile delinquents" (**168**, p. 310). *Result:* negative.

26. Landis reported that Heidbreder's Temperament Test (**123**, **124**) was of "much less value with juvenile delinquents" than with college groups (**168**, p. 311). *Result:* mainly negative.

27. Murray employed a personality test which she compiled largely from the Woodworth, Thurstone, and Laird tests and found that while it did not distinguish ordinary delinquents from non-delinquents, it did clearly differentiate from both these groups a specially selected experimental group of emotionally unstable male delinquents. She concluded: "86% of the experimental group score above the median of the control group, while only 10% of the control group score above the median of the experimental group" (**222**, p. 228). *Result:* positive.

28. Reusser gave the Sweet test to 423 delinquents in the Iowa Training School for Boys and to 419 normal school boys. He found significant differences between the two groups on five of its scales, but not on the Self-criticism and Feeling of Inferiority scales (**247**, p. 33). *Result:* mainly positive.

29. Slawson, giving the Woodworth-Mathews schedule to delinquent and normal groups of boys, found that "the delinquent boys as a whole are decidedly inferior in the status of their emotional stability as compared with a fairly unselected group of boys" (**282**, p. 234); and that "there is an intimate association between defective emotional makeup, as judged by the psychoneurotic responses made to the questions on this inventory, and male juvenile delinquency" (**281**, p. 279). *Result:* positive.

30. Smith prepared a questionnaire designed to measure inferiority feelings, gave it to normal children and juvenile court cases and reported recital ratios of 4.95, 2.41, 5.33, and 4.51 "in the direction of higher mean inferiority scores for Juvenile Court cases" (**285**, p. 633). *Result:* positive.

31. Snyder administered the Woodworth-Mathews schedule to delinquent groups of boys and girls. She found that they gave only slightly higher percentages of unfavorable responses to the test than did the normal children on whom it was standardized, and concluded that "this may or may not indicate slightly greater emotional instability among these delinquent children" (**287**, p. 183). *Result:* mainly negative.

32. Washburne gave an early edition of his Social-Adjustment Inventory

to 50 normal high school students and 50 prisoners of the same age range and found a critical ratio of 6.40 between the results for these groups (**332**, p. 129). *Result:* positive.

33. Wittman and Huffman gave the Bell Adjustment Inventory to normal high school students and to a group of teen-aged delinquents, and reported that "despite some rather marked differences on home adjustment and emotional adjustment . . . none of these differences are statistically significant. The wide range of scores for each group, from excellent to very poor adjustment in most categories, produced large standard deviations. There is a definite overlapping between the different groups" (**351**, pp. 178–179). *Result:* negative.

34. Wittman and Huffman administered the Guilford Martin Inventory (**105**, **106**, **107**, **108**, **109**, **191**) to the same groups of delinquents and normals and noted that "the greatest percentage of the controls and the delinquent boys are in the average and good adjustment ranges. Most of the delinquent girls are also in these ranges for health and social adjustment, but in home adjustment 44% are maladjusted and in the emotional 50% rate as poorly adjusted" (**351**, p. 180). *Result:* mainly negative.

In sum: Of 34 tests of the validity of personality questionnaires in the diagnosis of delinquency, 15 have proven to be positive or mainly positive, 6 questionably positive, and 13 negative or mainly negative.

VALIDATION BY PSYCHIATRIC OR PSYCHOLOGICAL DIAGNOSIS

The studies that may be found in the literature which report attempts to validate personality questionnaires by comparing scores of normal groups to which they have been given with those of psychiatrically or psychologically diagnosed neurotics, psychotics, or other abnormals include the following:

1. Anderson compared several case histories with Bernreuter test results. She gave no statistical evaluations, but concluded that "the objective measures . . . may render essential service to psychological counseling by individuals properly qualified professionally" (**11**, p. 469). *Result:* questionably positive.

2. Bennett (**19**) and Slater (**280**) reported on the administration of a neurotic inventory devised by E. T. O. Slater to 80 patients in a neurosis centre of the British armed forces and to 80 normal controls. Although this inventory contained only eight questions in its Anxiety scale, seven in its Hysteria scale, and eight in its Depression scale, critical ratios of 8.78, 4.60, and 8.82 respectively were obtained between mean scores of the neurotics and normals (**19**, p. 273). *Result:* positive.

3. Brown compared scores on his own Personal Inventory (**37**, **38**) for groups of 77 neurotic and 200 normal boys, and found a critical ratio of 14.56 (**38**, p. 89). *Result:* positive.

4. Darley had two counselors clinically interview 800 students and then gave them the Bell Adjustment Inventory. He found that it correctly diagnosed only 48.8% of the males and 35.5% of the females (**63**, p. 636). *Result:* mainly negative.

5. Darley utilized the Minnesota Inventory of Social Attitudes (**347**) on the same individuals, and found that it correctly diagnosed only 33.3% of the males and 40.0% of the females (**53**, p. 636). *Result:* mainly negative.

6. Darley and Ingle used the Bernreuter Inventory with 26 diagnosed psychotics and reported it to be ineffective in identifying their emotional maladjustments (**65**, p. 341). *Result:* negative.

7. De Angelis gave the Bernreuter schedule to 120 patients at the New York State Psychiatric Institute and concluded that "since only 40% of the psychiatric cases scored above the 84th percentile, the test as now constructed is not particularly valid in the light of actual human experience" (**67**, p. 76). *Result:* mainly negative.

8. Eysenck gave a composite personality questionnaire to hospitalized army neurotics and to normal army personnel. He found that it did not achieve "a high level of prediction" (**73**, p. 661). *Result:* mainly negative.

9. Feder and Baer administered the Bernreuter Inventory to 82 University of Iowa girls and compared their scores with a clinical evaluation of their personality and adjustment. They concluded: "A large amount of disagreement was found between the Inventory scores and the actual behavior records of maladjustment" (**75**, p. 143). *Result:* negative.

10. Feder and Mallet, using the Bell Adjustment Inventory on 21 diagnosed psychoneurotic and psychopathic college students and several classes of normal students, reported that "it appears that paper and pencil personality questionnaires of the types herein studied have doubtful validity for the discovery and diagnosis of personality maladjustments" (**76**, p. 15). *Result:* mainly negative.

11. Feder and Mallet found the same result for the Woodworth-House Inventory with the same groups (**76**, p. 15). *Result:* mainly negative.

12. Feder and Mallet reported the same result for the Thurstone Inventory used with the same groups (**76**, p. 15). *Result:* mainly negative.

13. Fisher and Watson gave their Inventory of Affective Tolerance (**77**, **78**, **340**, **341**, **342**) to 11 normal subjects and to 11 diagnosed neurotics. They reported: "The observed difference in scores between normal and neurotic subjects is statistically significant for both sexes" (**78**, p. 154). They also noted that their inventory did not differentiate between normal and psychotic groups, but claimed that it was not designed to do so (**78**, p. 156). *Result:* mainly positive.

14. Fisher and Hayes studied 416 normal Sarah Lawrence College students and 25 who had been psychiatrically diagnosed as maladjusted. They found that there were significant differences between the mean scores of the two groups on five of the six Bernreuter scales (**79**, p. 87). *Result:* positive.

15. Fleming gave the Woodworth P.D. Sheet to 341 Columbia college freshmen, and checked its results with their college activities, social life, sleep, illness, in college, and other measures of school and personal adjustment. They found no significant correlations, and concluded that "prediction of behavior, within the range of activities examined, is not possible from the scores obtained on any of the tests" (**82**, p. 56). *Result:* negative.

16. Fleming gave the Laird-Colgate Inventory (**127**, **165**, **166**, **167**, **294**) to 322 Columbia freshmen, and again found no significant correlation between their scores and their actual school and social life while at college (**82**, p. 56). *Result:* negative.

17. Garrett and Schneck studied 256 diagnosed psychoneurotic soldiers with the Woodworth P.D. Sheet and compared their scores with those of 103 normal Columbia College freshmen. They reported that "we cannot distinguish be-

tween the different neuroses nor between normals and abnormals on the basis of the gross P.D. scores alone" (**90**, p. 462). *Result:* negative.

18. Garrett and Schneck gave the Woodworth P.D. Sheet to Columbia College freshmen and found that most of the *high score or maladjusted* men as designated by the test were actually getting along as well as were the *normal* scorers (**90**, p. 460). *Result:* mainly negative.

19. Gilliland and Morgan gave the Neyman-Kohlstedt Extroversion-Introversion Test to 22 Dementia Praecox and 16 Manic-Depressive patients, and discovered that "the test almost completely separated the two groups with only one case of overlapping" (**95**, pp. 301–302). *Result:* positive.

20. Hanks compared the Thurstone-Clark scores of 92 freshmen at the University of Wisconsin with autobiographical records and materials from their college entrance application blanks and found a correlation of .57 between the two. He concluded: "There were indications that personality test scores were to some degree predictable" (**112**, p. 68). *Result:* questionably positive.

21. Hanna gave the Thurstone to 250 subjects applying for clinical assistance, previously dividing the group into adjusted and maladjusted subgroups. He reported that "as a group, those individuals considered sufficiently maladjusted as to require the services of a consulting psychologist or psychiatrist, showed far above average scores on the Personality Schedule" (**113**, p. 445). *Result:* positive.

22. Harriman administered the Woodworth-House Inventory to students who were entering college and later examined 47 of these students who turned out to be maladjusted. He discovered that "the scores were definitely predictive of subsequent maladjustment in not more than 30% of the cases. Many students considered maladjusted by college officials, faculty members, and other students did not indicate their problems in the inventory" (**114**, p. 259). *Result:* mainly negative.

23. Hathaway gave the Bell Adjustment Inventory to diagnosed psychopaths, and reported that "the Bell Adjustment Inventory has been unsatisfactory for this (diagnostic) purpose because it cuts off the normal scores too abruptly, not permitting extreme deviations in that direction" (**120**, p. 114). *Result:* negative.

24. Hathaway gave the Bernreuter Inventory to the same group of psychopaths and found that they characteristically obtained extremely normal or adjusted scores on the B1-N scale: so that, if the test were scored *backwards*, or its scores interpreted *opposite* to the manner advised in the instruction booklet, it would be a good diagnostic instrument for differentiating psychopathic inferiors from normals (**120**, pp. 114–117). *Result:* positive.

25. Hathaway found that, with the same group of constitutional psychopathic inferiors, the Humm-Wadsworth test also gave characteristically shaped profiles for the group, if they were interpreted in an inverse manner to that advised in the instruction booklet (**120**, pp. 114–117). *Result:* positive.

26. Hollingworth gave the Woodworth P.D. Sheet to 117 neurotic soldiers during the First World War and found that 52 of them received scores which are ordinarily considered normal on this test. Giving the inventory again to 202 neurotic soldiers after the armistice had been declared, he found that 151 of them received normal scores (**128**, p. 142). *Result:* negative.

27. House adapted the Woodworth P.D. Sheet for use with college students, and administered it to normal, psychoneurotic and psychotic males.

He reported that "no sharp lines of demarcation exist between the unselected normals and the declared psychoneurotics. No clear lines of cleavage separate the psychoneurotics from the psychotics. The 'normal curve of distribution' is perfectly applicable to the so-called abnormals" (**131**, p. 19). *Result:* negative.

28. Humm and Humm, validating the Humm-Wadsworth Temperament Scale against case studies taken before the tests were administered, found that "there is, in general, strong agreement between the scores . . . and the findings of carefully conducted case studies" (**135**, p. 63). *Result:* positive.

29. Ingles gave his own test of mental instability to 69 college students whom he had previously classified on the basis of the degree of mental instability judged to be present in each student. He found a coefficient of contingency of .78 for Form 1 and .81 for Form 2, "thus demonstrating a fairly high degree of validity as determined by this criterion" (**143**, p. 266). *Result:* positive.

30. Jackson administered the California Personality Test (**315**, **316**) to 100 high school students and also had the students interviewed, rated by the experimenter, and rated by their teachers and parents. He reported that "for five of the seven items of the entire self-adjustment schedule of traits utilized in this study, the group paper-pencil test yields the most effective evaluation" (**145**, p. 49). *Result:* mainly positive.

31. Jackson administered the Woody Student Inquiry blank to the same group and found that "the distribution of the effective evaluations by individual techniques discloses no particular pattern among the results. The interview procedure evaluates most effectively the traits of social adaptability, dominance, and phantasy, while teacher ratings contribute most effectively to the evaluations of the traits of inferiority and of seclusiveness. The traits of cooperativeness and of self-assurance are most effectively evaluated through the group paper-pencil testing methods" (**145**, p. 52). *Result:* questionably positive.

32. Johnson gave the Bernreuter Inventory to a group of subjects when they were presumably normal and when they were admittedly depressed and discovered that "the differences between the means for the two sets of Bernreuter scores are not large, but they are in the expected direction, suggesting that a depressed mood is accompanied by increased score for neurotic tendency and introversion" (**150**, p. 518). *Result:* questionably positive.

33. Landis, Ferrall and Page gave fear and anger questionnaires to a group of normal and abnormal subjects and concluded that "normal individuals report somewhat more intense fear and considerably more intense anger reactions than abnormal individuals" (**170**, p. 597). *Result:* mainly positive.

34. Landis, Zubin and Katz gave the Bernreuter Inventory to normal subjects and to neurotic and psychotic patients. They reported that it did not differentiate significantly between the two groups (**173**, p. 330). *Result:* negative.

35. Landis, Zubin and Katz, using the same groups, found that the Page Questionnaire of Schizophrenic Traits likewise did not significantly differentiate normals from neurotic and psychotic patients (**173**, p. 330). *Result:* negative.

36. Landis, Zubin and Katz discovered that the Maller Character Sketches did significantly distinguish the psychoneurotic group from a normal control group, but stated that "overlap between the two groups is still present" (**173**, p. 330). *Result:* mainly positive.

37. Landis and Katz gave the Bernreuter Inventory to 224 psychotic and

neurotic patients of the New York Psychiatric Institute and noted that 25% scored above the 90th percentile and 10% scored below the 10th percentile on the schedule. They concluded: "It has been variously reported that those receiving high scores are neurotic. Our study confirms this. However, the converse is not true. Those whose scores indicate that they are emotionally stable are not necessarily so. . . . If an individual receives a low score, no conclusions concerning his emotional adjustment to life may be drawn" (**172**, pp. 345, 354, 356). *Result:* mainly negative.

38. McCartney and Papurt gave a personal data questionnaire to prisoners at Elmira reformatory who were psychiatrically diagnosed and separated into normal and abnormal groups. They found a general tendency for normals to have lower pathological scores than other prisoners; but they reported no reliable statistical differences between the groups (**185**, p. 106). *Result:* questionably positive.

39. Maslow, Hirsh, Stein and Honigman gave a Security-Insecurity questionnaire to 122 college students and stated that it clearly differentiated well-adjusted from maladjusted subjects (**207**, p. 32). *Result:* positive.

40. Marsh tested the effectiveness of the Bell Adjustment Inventory in distinguishing between normal and seriously maladjusted college women. He discovered that while the differences between the two groups on the home adjustment scale of the inventory "are in the expected direction, they are probably not highly significant." This scale, moreover, was the most useful one in the inventory for predicting social and emotional maladjustment in the college women studied (**202**, pp. 105–108). *Result:* mainly negative.

41. Marshall used the Bernreuter schedule with 371 neuropsychiatric cases at the Lane Hospital of the Stanford Medical School and found that "50% of the psychoneurotics are more neurotic and more introverted than 80% of normals; 20% of the men and 10% of the women score higher than any case reported by Bernreuter. . . . The eight true schizophrenic males all rated above the 80th percentile on introversion, the 13 females rated about the 60th percentile; the four paranoid praecox cases of each sex, in contrast, all rated below the 50th percentile" (**205**, p. 601). *Result:* mainly positive.

42. Mathews gave the Woodworth-Mathews inventory to normal and maladjusted children and reported that "children selected because they show nervous or temperamental difficulties respond, as a whole, with more unfavorable answers than do those who are not so selected. . . . These findings are not so convincing as we would like" (**209**, p. 38). *Result:* questionably positive.

43. McNemar and Landis administered the Willoughby Emotional Maturity test (**349**, **350**) to a group of psychopathic women, and concluded that "there is no evidence that emotional maturity, of whatever the scale as used measures, is related to age, educational status, or clinical diagnosis" (**193**, p. 318). *Result:* negative.

44. Miles, Wilkins, Lester & Hutchens administered a Self-Descriptive Inventory to 2,543 successful and 921 failing Marine Corps recruits, and reported that "a score of 6 or over would select 80% of all failures (hits), 24% of all successes (false positives) and would miss 20% of all failures. The chi-square test at this score level indicates that so great a difference would arise less than one time in 100 by chance from similar random samples of the two groups if they were from a population actually homogeneous as to scores" (**216**, p. 256). *Result:* mainly positive.

45. Mittelmann, reporting on the administration of the Cornell Selectee Index (**360**) to 1200 army inductees, stated that "between 80 and 90% of those rejected or discharged are screened by the test. In achieving these figures about 10% of the healthy population is also screened" (**217**, p. 30). *Result:* mainly positive.

46. Mittelmann, writing about the application of the Cornell Selectee Index to 120 army hospital cases, 30% of whom were diagnosed neurotics, noted that "35% of the 120 patients were screened by the test. This number includes 87% of those found to have neuropsychiatric disturbances upon interview and a small number of subjects with mild neuroses and bad morale, as well as some *false positives*" (**217**, p. 30). *Result:* positive.

47. Morgan had an adaptation of the Thurstone Inventory administered to students entering college, and compared 146 well-adjusted subjects to 41 who were found to be neurotic by observation of friends, family, and personnel workers. He concluded: "This large group . . . showed no unusual or unique tendencies in either section which might indicate the effectiveness of the neurotic inventory as an instrument effective in pointing out neurotic individuals" (**219**, p. 184). *Result:* negative.

48. Neyman and Kohlstedt gave their Extroversion-Introversion test to 100 schizophrenics and 100 manic-depressive patients, and stated that "in a total of 200 patients examined there is 1% failures, 6% ill-defined reactions and 93% who gave scores exactly corresponding to their reaction types" (**226**, p. 486). *Result:* positive.

49. Page gave the Bernreuter Personality Inventory to 100 normal army trainees and to a matched group of 100 diagnosed psychoneurotics and found significant differences on the test between the two groups (**229**, p. 647). *Result:* positive.

50. Page, using the Psychoneurotic Inventory (which he compiled from the Bernreuter and the Psychosomatic Inventories), tested 500 undiagnosed army trainees, and found that "the Inventory had a coefficient of correlation with the psychiatric diagnosis of .78" (**229**, p. 657). *Result:* mainly positive.

51. Page, employing the Psychosomatic Inventory (**186**, **187**), again found significant differences between matched groups of normal army trainees and diagnosed psychoneurotics (**229**, p. 647). *Result:* positive.

52. Page, Landis and Katz gave a special questionnaire on schizophrenic traits to 100 manic-depressives, 125 schizophrenics, and 240 normals. They found that "the lower scores made by the manic-depressive patients as compared with the other two groups is completely reliable. The difference in mean score between the normal and schizophrenic group is not significant. . . . The normal and schizophrenic group possess on the average about the same number of schizophrenic traits" (**230**, p. 1217). *Result:* questionably positive.

53. Papurt administered the Woodworth P.D. Sheet to 50 female and 50 male patients at the Psychopathic ward of the Cleveland Hospital. He noted that "22% of the patients gave test results which according to Franz would classify these patients as decidedly normal; and yet each had been found insane by a staff of expert psychiatrists . . . 86% of the male patients had less than 30 pathological answers and 90% of the females had less than 30 pathological answers" (**233**, p. 344). *Result:* mainly negative.

54. Pedersen, giving the Bell Adjustment Inventory to 380 women freshmen at the University of Rochester, and gathering case materials on these

women, found that the health adjustment and social adjustment scales were fairly valid, but that the home adjustment and emotional adjustment scales were not. She concluded: "There is no real difference between the emotional scores of individuals rated emotionally maladjusted and other individuals" (**234**, pp. 229–236). *Result:* mainly negative.

55. Pescor experimented with 800 inmates of the U. S. Northeastern Penitentiary Hospital by administering the Woodworth P.D. Sheet. He found that the emotionally unstable group, as diagnosed by the Woodworth test, "showed greater inability to make a satisfactory adjustment as exemplified by more frequent attendance on sick line, more admissions to the hospital, more assignment to quarters and convalescent detail because of illness and more frequent subjection to diciplinary measures than the control group. . . . The findings on the Woodworth P.D. Sheet were correct in about 85% of the cases as revealed by subsequent observation" (**237**, p. 393). *Result:* mainly positive.

56. Poe Eng Yu, upon administering the Bernreuter P.I. schedule to manic-depressive and schizophrenic patients, found that it differentiated the manic-depressives from the schizophrenics only slightly, and "not sufficiently to warrant its use as a diagnostic tool." It did, however, differentiate normals from manic-depressives and schizophrenics (**243**, p. 587). *Result:* questionably positive.

57. Reed and Wittman gave the Humm-Wadsworth test to 417 psychiatric patients at the Elgin State Hospital and to a control group of 217 normals. They found that while the Normal and Cycloid components of the test differentiated controls from experimental group, the Schizoid, Hysteroid, and Epileptoid scales did not. They concluded that "Normal and Cycloid components coincide with psychiatric diagnosis. Other components do not" (**245**, p. 592). *Result:* questionably positive.

58. Roberts and Fisher gave the Bernreuter Inventory to 35 students at New York University after a trained interviewer had previously diagnosed each student by means of a 30 to 60-minute interview. They obtained a correlation of .317 between interviewer's diagnoses and Bernreuter scores, and noted that "this correlation is too low to be significant" (**252**, p. 92). *Result:* mainly negative.

59. Rosenwald administered the California Personality Scale to schizophrenic and manic depressive patients and to an adequate control group. He reported that "it failed to differentiate the psychotic groups or show significant differences between experimental and control groups. The low positive correlation with an apparently valid prognostic scale indicates its undesirability as an aid in psychiatric diagnosis" (**358**, p. 599). *Result:* negative.

60. Runner and Seaver gave a personality analysis test to 73 poorly adjusted and 87 well-adjusted adults, and stated that "a significant ratio of difference of 13.75 was found between the two groups" (**267**, p. 222). *Result:* positive.

61. St. Clair administered the Bernreuter scales to 729 male and 433 female freshmen at Temple University. He checked their scores on the Bernreuter against information obtained from these students on other questionnaires which asked them about health, home conditions, personal problems, previous interests and activities, and certain social attitudes. His conclusion was: "The Bernreuter scale appears to have a degree of validity as a measure of personality traits" (**270**, p. 531). *Result:* mainly positive.

62. Sargent, in a footnote to her study of projection in paper and pencil,

personality tests, reported that "in a recent reading of state hospital records the writer noted that a large majority gained scores of *Excellent Adjustment* on a well-known inventory which has since been abandoned by the institution for this reason" (**271**, p. 1). *Result:* negative.

63. Schott administered the Thurstone test to 300 neuropsychiatrically diagnosed patients and found that while there was "a marked difference in the distribution when the scores of our patients are compared with those of Thurstone's college subjects, and with those of 200 normal subjects . . . the test failed to indicate the degree of maladjustment in individual cases" (**273**, p. 239). *Result:* mainly positive.

64. Seibert gave the Johnson Temperament Analysis test to 154 incoming women students in Green Mountain Junior College, and commented on the significance of the fact that of the 21 students with measured indications of maladjustment on this test who later presented themselves for counseling, confirmation of their maladjustment was found in each case. He concluded: "Experience of one year warrants the further use of this device in a battery of tests on the basis of its value as an aid in the detection of women students with undesirable and ineffective traits of temperament" (**275**, p. 194). *Result:* positive.

65. Stagner used the Bernreuter test with a group of college students, and then interviewed those students who came to his office to get the results of their tests. He stated, on the basis of this check-up, that "the validities of the scales B1-N and B2-S of the Bernreuter are high" (**295**, p. 418). *Result:* positive.

66. Stogdill and Thomas gave the Bernreuter inventory to well-adjusted and maladjusted college students and found that "as a measure of adjustment the Bernreuter Personality Inventory appears to be very helpful in discriminating between well-adjusted and maladjusted students" (**297**, p. 313). *Result:* positive.

67. Sweet gave his own personality test to 62 institutionalized delinquents on whom intensive case studies were made. He found that the case studies agreed with personality test scores 32 times, agreed partially in 11 instances, and disagreed in 14 cases (**300**, p. 33). *Result:* questionably positive.

68. Travis examined ten subjects with his own personality questionnaire after obtaining extensive case histories on them, and found a 60% to 95% agreement between the questionnaire results and the histories. "There was a rather high relationship," he claimed, "between the individual's characteristic reactions to the test and his mental conflicts" (**317**, p. 418). *Result:* mainly positive.

69. Turney and Collins gave the Bernreuter test to 21 high school students in Leroy, Kansas, and reported, without any statistical confirmation, that "the test results were corroborated by case studies of both experimental and control groups. These case studies indicate that the test results are valid" (**324**, p. 553). *Result:* mainly positive.

70. Turney and Collins made the same report for the use of the Thurstone test with the same group (**324**, p. 553). *Result:* mainly positive.

71. Turney and Collins made the same report for the use of the Maller test with the same group (**324**, p. 553). *Result:* mainly positive.

72. Warner and Gallico administered the Cornell Service Index to 242 naval neuropsychiatric patients and stated that "the results obtained distinguish with a high degree of accuracy patients without apparent personality disturbances from those who present psychiatric complaints of significant degree.

The form does not effect a very clear separation between mild and severe personality disturbances, though it is of some help in this differentiation. The form does not reveal adequately the histories of subjects who are lacking in awareness of their difficulties. It does not cover many cases of conversion hysteria, and it does not concern itself with sexual disturbances" (**331**, p. 217). *Result:* mainly positive.

73. Watson gave the Watson-Fisher Test of Affective Tolerance to college students who came voluntarily to the Student Personnel Bureau of the City College of New York for help in regard to some problems, and found that "male students, who are found to need assistance in regard to affective problems through the medium of a Student Personnel Bureau . . . score significantly lower on the 'Inventory of Affective Tolerance' than those needing assistance in regard to other matters . . . . It would appear, then, emotional maladjustment or, as the writer would prefer to call it, lack of affective tolerance, can be identified in male college students with considerable success by use of this inventory" (**341**, pp. 13–14). *Result:* positive.

74. Weider, Brodman, Mittelman, Wechsler and Wolff, reporting on the administration of the Cornell Index to 970 enlisted army and navy men, 539 of whom were normal controls and the rest diagnosed psychoneurotics, concluded that "by means of the Index score a sharp differentiation can be made between persons with little or no psychoneurosis and those in whom psychoneurosis interferes or may interfere seriously with military performance. Ninety-eight per cent of severe psychoneurotics are detected while 3.5 per cent of ostensibly healthy persons are found to have poor scores" (**343**, p. 5). *Result:* positive.

75. Wittman and Huffman gave the Bell Adjustment Inventory to 191 normal high school students and to 223 psychotics of the same age range and found that the scores of the psychotics did not differ reliably from those of the normals (**351**, pp. 178–179). *Result:* negative.

In sum: Of 75 attempts to validate personality questionnaires against psychiatric or psychological diagnoses, 36 have shown positive, 9 questionably positive, and 30 negative results.

VALIDITY BY RATING DIAGNOSIS

There have been a good many reports in the literature of attempts to validate personality questionnaires by checking their results against respondents' ratings by others who presumably know them pretty well: for example, by the subjects' teachers, friends, or associates. These include the following studies:

1. Allen administered Flugel's personality questionnaire to 60 students from University College, London, and had them rated by two friends. Correlation between the questionnaire scores and the ratings of the first friend was .49, and between scores and the ratings of the second friend was .41 (**5**, p. 41). *Result:* questionably positive.

2. Allport validated his original Ascendance-Submission questionnaire (**7, 16, 36, 264, 330**) against associates' ratings of his subjects and found a correlation coefficient of .459 between the two. He admitted, however, that "inasmuch as the same ratings were employed to secure the score values and to obtain

these measures of validation, the correlations are probably too high" (**7**, p. 133). *Result:* questionably positive.

3. Bernreuter gave his own Self-Sufficiency Test (which was later incorporated into his Personality Inventory) to 388 college men and 456 college women, and noted that "all of the ratings (of their associates) correlate positively with scores in the SS test." The final correlation between associates' ratings and SS test scores, however, was only .44 (**27**, p. 299). *Result*: questionably positive.

4. Broom gave the Laird-Colgate Inventory to college juniors and seniors and had them rated by their associates. His conclusion was that "in terms of the criterion set up, the findings of the rating scale, the test has a satisfactory degre of validity, since the coefficients are indicative of marked correlation" (**35**, p. 123). *Result:* positive.

5. Burks reported on the administration of the Bernreuter Inventory to 50 freshman girls at Mills College. Eight pooled ratings by hall officers and the head resident of these students were correlated with the test scores, the correlation coefficient for the B1-N scale being .21x (**39**, p. 148). *Result:* mainly negative.

6. Cady gave the Woodworth-Cady test to 150 boys whom he had rated, by their teachers, on incorrigibility. He stated that "correction for attenuation failed to yield coefficients in excess of .55. We must conclude, therefore, that these tests do not measure incorrigibility as indicated by the criterion used, although the latter has a very high reliability" (**44**, p. 120). *Result:* mainly negative.

7. Campbell gave a questionnaire employing the Freyd Intro-Extroversion list to insane subjects who were also rated by physicians associated with them, and reported that "the correlations between the ratings of self and associates was .37" (**45**, p. 481). *Result:* mainly negative.

8. Casselberry had 70 office workers rated by their associates, and then administered the Bernreuter Inventory to them. He noted a correlation coefficient of .47 between the test scores and the ratings, and concluded that "this would seem to indicate a certain degree of ability on the part of the Inventory to measure personality" (**49**, p. 583). *Result:* questionably positive.

9. Clark and Smith had 138 students at the Rochester Atheneum and Mechanics Institute rated by faculty members, and then had these ratings checked by school counselors. They found that the correlation between the ratings and the students' scores on the Bell Adjustment Inventory was *quite low* (**52**, p. 84). *Result:* mainly negative.

10. Clark and Smith reported the same result for the same group with the Washburne Social Adjustment Inventory. They concluded that "all of the positive correlations are so low as to invalidate them for individual prognostic purposes" (**52**, p. 86). *Result:* negative.

11. Cook gave the Bernreuter Inventory to 100 subjects who were then rated by four independent estimators who knew them well. He stated that "in eight cases out of one hundred, there was a marked difference (approximately 50%) between the two evaluations, and in each case this was due to a relatively low estimate of neurotic tendencies on the Bernreuter scale in conjunction with high B2-S and B4-D scores" (**57**, p. 159). *Result:* mainly positive.

12. Drake, Roslow and Bennett gave the Link P.Q. Inventory to 70 high school boys and 44 girls and had them rated by their peers. They found

"significant positive correlation exists between four scales of the P.Q. test and the judgment of classmates" (**68**, p. 213). *Result:* positive.

13. Flemming and Flemming gave the Woodworth-Mathews test to 88 girls at the Horace Mann High School, New York City, and discovered that "the correlation of −.148 between teachers' estimates of emotional balance and the Mathews revision of the Woodworth P.D. questionnaire shows no relation between this supposed test of emotional stability and the criterion of teacher's estimates of emotional balance as defined in this study . . . . It is necessary to conclude that the test is not a valid test of emotional balance" (**84**, p. 505). *Result:* negative.

14. Guilford and Martin gave their own personality test to University of Southern California students and had five friends of each student rate their subjects. They reported that "evidence from self-ratings and ratings by close associates shows that the scores for Factors S, D, and R are quite valid. Ratings for Factors T and C were not themselves sufficiently acceptable as criteria against which to validate scores for those factors" (**109**, p. 229). *Result:* mainly positive.

15. Hunt gave the George Washington Social Intelligence Test to 98 sales employees who were rated by their executive superiors, and found a correlation of .61 between the ratings and the test scores. She gave the same test to a group of college students who had been rated by their teachers and by sorority and fraternity associates and found correlations between the ratings and the scores averaging about .40 (**140**, p. 324). *Result:* questionably positive.

16. Jarvis and Johns gave the Bernreuter Inventory to 148 students at the Rochester Atheneum and Mechanics Institute and had them rated by several counselors who were well-acquainted with the students. They reported: "The relation between the scores made on the Bernreuter Inventory and ratings given by counselors varied markedly from counselor to counselor, with the coefficients of correlation ranging from −.15 to .14 . . . . The Bernreuter Personality Inventory offers little advantage in the isolation of personality problems peculiar to our particular educational situation" (**147**, p. 9). *Result:* negative.

17. Jasper gave his own Depression-Elation test to 34 College seniors and had them ranked by six professors and by their class officers. He obtained a correlation of .71 between the ratings and the test scores, and noted that "when the coefficient of validity obtained from the pooled judgments is corrected for attenuation it is raised to $r=.95$" (**148**, p. 315). *Result:* positive.

18. Keys and Guilford examined 197 boys and 208 girls in Chico, California high schools with the Bernreuter Inventory, and also had the students rated by their homeroom teachers and by their principal, counselors, and dean. They found that the correlation between the test scores and the ratings were not "sufficient for accurate prediction of the behavior of individuals pupils" (**153**, p. 652). *Result:* mainly negative.

19. Keys and Guilford, using the Bell Adjustment Inventory on the same group, found the same negative results (**153**, p. 652). *Result:* mainly negative.

20. Keys and Guilford, using the Loofbourow-Keys Inventory with the same groups, found only slightly better results, and reported that the scores "must be regarded as suggestive only" (**153**, p. 652). *Result:* mainly negative.

21. Keys and Whiteside administered the Woodworth-Cady test to 182 grade school children who had been rated by their teachers. They concluded:

"Children's scores on the Woodworth-Cady questionnaire of Psychological Symptoms correlate well (in this study .53±.036) with an average of teachers ratings of these children on nervous and emotional traits" (**154**, p. 441). *Result:* questionably positive.

22. Landis gave the Thurstone test to 112 girls committed to Long Lane Farm and checked their scores against periodical character ratings by matrons in charge of the girls. He stated that "it seems only fair to conclude that either the tests which we used did not bear any essential correspondence to the character traits as they were formulated by the institution authorities or that the character traits had not afforded well-defined entities which are open for experimental classifications" (**168**, p. 301). *Result:* negative.

23. Landis, Gullette and Jacobsen administered the Woodworth P.D. Sheet to 12 men, 12 women and one boy, and had these subjects rated by their associates. They found a correlation of .31 between the ratings and the test scores. (**171**, p. 232). *Result:* mainly negative.

24. Marsh had the hall counselors at Stephens College rate first-year women and divide them into *successful* and *unsuccessful* groupings. He found that when the Washburne Social Adjustment Inventory was given to these college women, "neither total adjustment scores, nor sub-test adjustment scores can be used to predict the behavior of first-year college women" (**203**, p. 293). *Result:* negative.

25. McKinney had fraternity brothers rate each other and then gave the Thurstone Personality Schedule to groups of well- and poorly-adjusted students selected by these ratings. He noted that "the scores on the Thurstone Personality Schedule show a statistically reliable difference between the two groups" (**192**, p. 442). *Result:* positive.

26. Naccarti and Garrett gave the Woodworth P.D. Sheet to 54 Columbia College students who were rated by their instructors at the beginning and end of a school term. They found that the group of students designated as neurotic was "considerably lower in the average rating for emotional stability, both in self-estimate and instructor's estimate. This result would tend to increase the probability of real differences in emotional stability, which has been strongly suggested by the results from the Woodworth P.D. Sheet" (**224**, pp. 261–262). *Result:* positive.

27. Oliver had subjects rate themselves on an extro-introversion questionnaire and then had their friends rate them on the same traits. He concluded that "subjects who would be classed as extroverts or introverts by their own ratings would by no means necessarily be so classed by the ratings of their friends" (**227**, p. 351). *Result:* mainly negative.

28. Pallister gave Lecky's Individuality Record to 200 Barnard women and also had them rated by three of their classmates. She reported that "the validity of the questionnaire is roughly indicated by the general higher average WA (withdrawal) score of those whom associates rate as falling on the inferior side than of those falling on the superior side of the average on a number of rating scales which deal with characteristics studied in the various categories of the Individuality Record." However, for 108 subjects, her tables show critical ratios of only 1.0056 for the trait of "nervousness" and .9931 for "temperament" (**232**, pp. 53–4). *Result:* mainly negative.

29. Peters administered Bernreuter, Bell, and Link P.Q. Inventories to groups of subjects and had these checked against Guess Who ratings by their associates. He concluded: "The validity correlations ranged from −.078 to

.503 and averaged about .26. All but two of them were highly significant statistically" (**238**, p. 453). By "highly significant statistically," he probably means significantly different from zero. *Result:* mainly negative.,

30. Pintner and Forlano gave their Aspects of Personality Inventory (**240, 241, 242**) to 582 grade school children, and had the children's teachers rate them on their outstanding personality characteristics. Comparing the test scores to the teacher-rated groupings of the children, they stated that "the differences between the means were in general rather small, but the general trend of the difference seemed to be in the expected direction" (**240**, pp. 30–1). *Result:* mainly negative.

31. Reed administered the Bernreuter Inventory to 259 Freshmen entering Fort Hays Kansas State College and had their teachers later rate them on the same traits covered by this test. He concluded that "the Bernreuter scores have little or no relationship with ratings of traits by the same men as they are given by teachers well acquainted with the pupils rated" (**244**, p. 435). *Result:* negative.

32. Root gave his own Introversion-Extraversion Test (**256**) to 78 subjects, subjects, and had them rated on introversion-extroversion by two psychology instructors and one graduate student who were acquainted with them. He reported a decided difference in the scores of the five respective groups into which the raters divided these subjects (**255**, p. 252). *Result:* positive.

33. Root and Root gave the Neyman-Kohlstedt Diagnostic Test to students who were rated by a jury of two psychologists and two graduate students on a five-point scale for introversion-extroversion. They found that "this study of the Neyman-Kohlstedt test tends to corroborate the findings of the designers" (**257**, p. 421). *Result:* mainly positive.

34. Roslow gave the Link P.Q. test to 34 girls in a private school, and reported that when the rank order coefficients between their scores and their teachers' ratings on corresponding questions were examined, they were found to "range from .33 to .74 and indicate marked agreement between the P.Q. test results and the teachers' composite answers" (**260**, p. 539). *Result:* questionably positive.

35. Schwegler gave a questionnaire consisting of 31 items taken mainly from the Woodworth P.D. Sheet to 206 individuals, half of whom had been rated as introverts and half of whom had been rated as extroverts by several of their teachers. He found a critical ratio of 6.3 between the mean scores of the two groups on this test (**274**, p. 138). *Result:* positive.

36. Smith gave his own Inventory for the Measurement of Inferiority Feelings to 154 high school students and also had them rated by their teachers, class advisers, and two friends. He stated that "the correlations are all low" (**286**, p. 83). *Result:* negative.

37. Thompson gave the Link P.Q. test to 146 boys and 113 girls attending junior high school at Mooseheart, Illinois, and reported that "with the exception of economic self-determination, sub-trait and extroversion scores were found to be in essential agreement with ratings by the Dean of Men in selecting boys with the greatest and least amount of each trait. Similar ratings by the Dean of Women for the girls did not agree so well with test scores" (**310**, p. 288). *Result:* questionably positive.

38. Traxler administered the Bell Adjustment Inventory to 33 high school students whom he had rated by from three to eight teachers on the traits it endeavors to score. He found that the correlations between its emotional

adjustment scale and the teachers' ratings was .318; and the correlation between the total Bell scores and the ratings was .463. He concluded that "the low correlations of the scores with teacher judgment fail to substantiate the validity of the inventory" (**319**, p. 677). *Result:* mainly negative.

39. Travis employed his own diagnostic questionnaire on three well-known individuals whom he had rated by 20 of their intimate associates, and found correlations of .49, .24, and .07 between the test scores and the ratings. He then had eight teachers rate 63 pupils to whom he had given the test, and reported a correlation of .19 between their ratings and the pupils' test scores (**317**, p. 412). *Result:* mainly negative.

40. Turney and Fee gave the Bell Adjustment Inventory to 85 pupils at the University of Kansas and had them rated by 15 teachers and supervisors. The reliability of the ratings themselves was high, ranging from .757 to .869; but the correlations between the ratings and the Bell scores were only .279 and .215. They note that "if the ratings are highly valid then it follows that the Adjustment Inventory does not possess a high degree of validity" (**325**, p. 196). *Result:* mainly negative.

41. Washburne gave his OSPA Social Adjustment test to 74 matched public schools boys and girls whom teachers and principals had rated as having exceptionally good or exceptionally poor adjustment, and found a critical ratio of the difference between the mean scores of the two groups of 8.57 (**332**, p. 129). *Result:* positive.

42. Wasson checked college students' scores on the Wilboughby Emotional Maturity scale against outside ratings of these students and found correlations between .60 to .70, which he considered to demonstrate "rather low" validity (**334**, p. 362). *Result:* questionably positive.

43. Whitman gave a modified Laird-Colgate Inventory to 80 students in an elementary class in psychology, and had the students rated by two of their friends. He reported that "the results for validity give a correlation of .576 with a P.E. of .054" (**344**, p. 503). *Result:* questionably positive.

44. Wrightstone gave the Woodworth-Mathews P.D. sheet to 92 Junior high school students and had their scores compared with case studies by visiting-teachers and with systematic observations by other teachers. He gave no statistical data for the test as a whole, but stated that in items such as social adjustment, home conditions, nervous habits, daydreaming, fear of things, persecution complex, and phobias or manias, there was an agreement of 91.3% between independent analyses by the investigator and descriptions by the school staff (**354**, p. 44). *Result:* positive.

In sum: Of 44 attempts to validate personality questionnaires against the ratings given to respondents by their teachers, friends, or associates, results have been positive in 12 cases, questionably positive in 10, and negative in 22.

VALIDATION ACCORDING TO TEST INTERCORRELATIONS

Paper and pencil personality questionnaires have largely been based on questions taken from previously constructed schedules, and some of them have been frankly validated by their authors against the tests on which they were based. While this method of validation is, to say the

TABLE I

INTERCORRELATIONS OF PERSONALITY QUESTIONNAIRES

Investigators	*Questionnaires Investigated*	*Intercorrelations Found*	*Indication of Validity*
Adams (**1**, p. 147)	Adams-Lepley and Bernreuter	.00 to .42	Mainly negat
Bernreuter (**25**, p. 384)	Bernreuter and Thurstone	.84 to 1.00	Positive
Bernreuter (**25**, p. 384)	Bernreuter and Laird	.84 to 1.00	Positive
Bernreuter (**26**, p. 195)	Laird and Root	.384	Mainly nega
Bernreuter (**26**, p. 195)	Heidbreder and Root	.346	Mainly nega
Bernreuter (**26**, p. 195)	Heidbreder and Laird	.444	Questionably positive
Bernreuter (**26**, p. 195)	Root and Neyman-Kohlstedt	.602	Questionably positive
Bernreuter (**26**, p. 195)	Laird and Neyman-Kohlstedt	.248	Mainly negative
Bernreuter (**26**, p. 195)	Heidbreder and Neyman-Kohlstedt	.280	Mainly negative
Bernreuter (**26**, p. 195)	Willoughby and Root	.369	Mainly negative
Bernreuter (**26**, p. 195)	Willoughby and Laird	.422	Questionably positive
Bernreuter (**26**, p. 195)	Willoughby and Heidbreder	.478	Questionably positive
Bernreuter (**26**, p. 195)	Willoughby and Neyman-Kohlstedt	.220	Mainly negative
Bernreuter (**26**, p. 195)	Root and Bernreuter	.298	Mainly negative
Bernreuter (**26**, p. 195)	Laird and Bernreuter	.382	Mainly negative
Bernreuter (**26**, p. 195)	Heidbreder and Bernreuter	.514	Questionably positive
Bernreuter (**26**, p. 195)	Neyman-Kohlstedt and Bernreuter	.162	Negative
Bernreuter (**26**, p. 195)	Willoughby and Bernreuter	.646	Questionably positive
Boynton and Walsworth (**32**, p. 90)	Maller and B.P.C.	.53	Questionably positive
Boynton and Walsworth (**32**, p. 90)	Maller and Bernreuter	.67	Questionably positive
Boynton and Walsworth (**32**, p. 90)	Maller and Rogers	.12	Negative
Boynton and Walsworth (**32**, p. 90)	Bernreuter and B.P.C.	.65	Questionably positive
Boynton and Walsworth (**32**, p. 90)	Bernreuter and Rogers	.10	Negative
Boynton and Walsworth (**32**, p. 90)	B.P.C. and Rogers	.33	Mainly negative
Clark and Smith (**52**, p. 88)	Bell and Washburne	.288	Mainly negative
Gilliland (**94**, p. 409)	Bernreuter and Laird	.47	Questionably positive
Gilliland (**94**, p. 409)	Bernreuter and Marston	.37	Mainly negative

TABLE I—*(continued)*

Investigators	*Questionnaires Investigated*	*Intercorrelations Found*	*Indication of Validity*
Gilliland (**94**, p. 409)	Bernreuter and Northwestern I-E	−.09	Negative
Gilliland (**94**, p. 409)	Laird and Marston	.30	Mainly negative
Gilliland (**94**, p. 409)	Laird and Northwestern	.10	Negative
Gilliland (**94**, p. 409)	Marston and Northwestern	.25	Mainly negative
Greene and Staton (**99**, p. 655)	Bernreuter and Willoughby	.37	Mainly negative
Greene & Staton (**99**, p. 655)	Bernreuter and Bell	−.62	Questionably positive
Greene & Staton (**99**, p. 655)	Bell and Willoughby	.55	Questionably positive
Jackson (**144**, p. 197)	California and Woody	−.38 to .50	Mainly negative
Jasper (**148**, p. 315)	Jackson Depression-Elation and Conklin	low	Negative
Page (**229**, p. 649)	Bernreuter and Psychosomatic	.63	Questionably positive
Perry (**236**, p. 93)	Bernreuter and Laird	.632	Questionably positive
Pintner (**239**, p. 766)	Bernreuter and Thurstone	.79	Mainly positive
Root (**255**, p. 252)	Root and Neyman-Kohlstedt	.831	Positive
Smith (**285**, p. 83)	Smith Inferiority Feeling and Bernreuter	.60	Questionably positive
Sperling (**292**, p. 297)	Sperling Human Behavior and Bell	.736	Mainly positive
Sperling (**292**, p. 297)	Sperling Human Behavior and Thurstone	.748	Mainly positive
Sperling (**292**, p. 297)	Bell and Thurstone	.785	Mainly positive
Stagner (**293**, p. 462)	Neyman-Kohlsted and Laird	−.392	Mainly negative
Stagner (**293**, p. 462)	Thurstone and Laird	.530	Questionably positive
Stagner (**293**, p. 462)	Thurstone and Neyman-Kohlstedt	−.340	Mainly negative
Tubbs (**323**, p. 350)	Adams-Lepley and Bernreuter	low	Negative
Wasson (**334**, p. 362)	Willoughby and Humm-Wadsworth	high	Positive
Watson (**340**, p. 85)	Watson-Fisher and Bernreuter	.66	Questionably positive
Watson (**340**, p. 85)	Watson-Fisher and Willoughby	.70	Mainly positive
Watson (**340**, p. 85)	Watson-Fisher and Bell	.56	Questionably positive
Watson (**340**, p. 85)	Watson-Fisher and Laird	.30	Mainly negative
Watson and Fisher (**342**, p. 148)	Watson-Fisher Affective Potency and Bernreuter	low	Negative
Watson and Fisher (**342**, p. 148)	Watson-Fisher Affective Potency and Willoughby	low	Negative

least, highly suspect, it is nonetheless true that intercorrelations between personality inventories may afford us an inferential estimate of their validity. If two tests, particularly questionnaires which have a similar format and content, are designed to measure the same trait, and they correlate highly with each other when used with the same groups, we have no adequate proof that they actually measure that trait; but we do know that they measure *something* in a consistent manner. Similarly, if the same two tests have a very low intercorrelation, we may justifiably infer that at least one of them does *not* measure the trait in question (while the other one may or may not measure it accurately).

A review of the experiments which have brought forth intercorrelation figures for various personality questionnaires may consequently add to our knowledge of the general validity to be expected from such schedules. Making such a review, we find 55 reported studies, which we list in Table I.

It can be seen from an examination of Table I that of 55 attempts to intercorrelate similar personality questionnaires, 9 gave positive, 18 questionably positive, and 28 negative indications of validity.

VALIDATION ACCORDING TO OVER-RATING EXPERIMENTS

Some of the most interesting experiments that have a bearing on the question of personality test validity have been those specifically concerned with discovering whether or not respondents answer self-descriptive questionnaires with a high degree of accuracy and truthfulness. For if it be conceded that testees habitually over-rate or lie about themselves on self-marking instruments, definite suspicion is cast on the validity of personality schedules that do not make some special allowances for this over-estimating tendency.

Attempts to evaluate the truthfulness of respondents to personality questionnaires have been reported as follows in the literature:

1. Allport found significant differences in scores on the Thurstone Personality Inventory between students at Dartmouth and the University of Chicago. He felt that these differences "can best be explained, it seems, by the attitudes of the students. The Chicago subjects, being freshmen, were probably more on the defensive, less given to self-revelation, than the psychologically sophisticated and test-hardened Dartmouth students" (8, p. 527). *Result:* probably negative.

2. Bain gave the same questionnaire material to a number of subjects on two different occasions, and found that one-quarter of the total number of items given were changed between the two testings. Questions asking for subjective personal information about the respondents were changed with significantly greater frequency than were those merely asking for factual data (14, p. 445). *Result:* mainly negative.

3. Bell stated that his experience with his own Adjustment Inventory led him to the conclusion that better results will be obtained if the test is given in small groups by the person who will later counsel the respondents; if special efforts are made to enlist the cooperation of the subjects; and if the students who take it "do not sit too near to each other" (**18**, p. 31). *Result:* mainly negative.

4. Bennett and Gordon discovered a tendency of nursing students "to modify their responses in a favorable direction" when they were given the Bernreuter as part of the selection process of their being enrolled in a nursing school (**20**, p. 277). *Result:* negative.

5. Bennett and Gordon found the same over-rating tendency present when the Minnesota Personality Test (**65**, **347**) was employed with nursing students (**20**, p. 277). *Result:* negative.

6. Bennett administered the Multiphasic Test to ten homosexuals and found that they gave positive results on the feminine scale when assured that they would not be punished for their answers; but when told to conceal their homosexuality if they could, six of the nine gave negative results to the same scale (**22**, p. 417). *Result:* negative.

7. Bernreuter gave his own Personality Inventory to a group of college students who first answered it in the regular manner, and were then asked to answer it "in the manner which they believed would result in their gaining the greatest possible degree of social approval." He found correlations of only −.07 between the two answerings on the B1-N scale. He similarly found only a slight relationship (a correlation of .22) between what an individual reported himself to and what he would like to be on the B1-N scale (**25**, p. 385). *Result:* positive.

8. Bonney gave a questionnaire to grade school pupils asking about their personal doings and discovered that only 27% of their answers were completely accurate, while 43% were approximately accurate. He noted that "the pupils erred by over-estimating rather than by under-estimating those behaviors which are the object of considerable moral pressure or which affect their social status" (**31**, p. 119). *Result:* negative.

9. Cogan, Conklin and Hollingworth had college women rate themselves and their classmates on various traits, and concluded: "In all cases the individual places herself further from her true position than do her friends on the average. In general the error of self-estimation tends to be half again as great as the average error of the judgments of associates. . . . In the case of *undesirable* traits this constant error is toward underestimation" (**53**, p. 174). *Result:* negative.

10. Copeland studied the reliability of adults' answers to questions about their height, weight, and age, and found that "even these simple items of personal history are not sufficiently reliable for purposes of individual diagnosis" (**58**, p. 251). *Result:* negative.

11. Crook made a special analysis of the distributions of responses among the five response categories on the Willoughby Personality Schedule given to several populations. He concluded that "evidence is offered that on certain items the shape of the distribution is determined largely by a systematic tendency for the subjects to give themselves more favorable ratings than they deserve" (**60**, p. 194). *Result:* negative.

12. Dudycha gave a questionnaire on dependability to students who were already known to be dependable or undependable. He discovered that "dependable students do not answer the questions in a favorable manner and undependable ones in an unfavorable manner, but . . . both tend to mark the favorable answers, except that the undependable students, in some question at least, mark the less favorable answers. Then again, we have observed that the undependable group answers some of the questions more favorably than the dependable ones do" (69, p. 52). *Result:* negative.

13. Gordon and Davidoff gave the Washburne Inventory to students in four Philadelphia junior high schools, and made certain that special instructions about answering the tests truthfully were given at the time of their administration. They nevertheless found that 15% of the questionnaires had lie scores so high that they had to be discarded (96, p. 54). *Result:* mainly negative.

14. Hartmann found that after taking a course in educational psychology, the scores of students on the Bernreuter Inventory rose significantly (118). This might be interpreted as resulting from the students' increased sophistication. *Result:* probably negative.

15. Heidbreder had subjects rate themselves on personality traits and then later tell how they would like to possess the same traits. She reported that "in this study there are large and significant differences between self-ratings in personality traits and ratings indicating preferences for traits. The differences occur whether preferences refer to the presence of the trait in the self or in others. The self-ratings, therefore, may be regarded as constituting a very different reaction from that involved in ratings which avowedly indicate preferences" (125, p. 74). *Result:* positive.

16. Hollingworth found that when the Woodworth P.D. Sheet was given to neurotic soldiers in the first World War before and after the armistice was signed, there was a "striking difference between positive symptoms reported by over two-thirds before and by less than one-sixth after the armistice." He interprets this result as proving that while the war was still going on, the soldiers under-rated themselves on the Woodworth while after it had ended they over-rated themselves in an effort to get an early release from the army (128, pp. 193, 194). *Result:* negative.

17. Humm, discussing the Humm-Wadsworth Temperament Scale, admitted that "frankness of response has been found to vary somewhat with the test conditions," and that normally 10 to 20% of the tests have to be rejected because the lie score is too high (134, p. 529). *Result:* mainly negative.

18. Humm and Humm, making a special study of the No-count scores on the Humm-Wadsworth scale, concluded that "when cases whose No-count is outside the acceptable range are considered, it is observed that subjects with too high No-counts tend to over-report Normal and under-report the other components (except Epileptoid), while cases with too-low No-count tend to under-report Normal and over-report the other components (except Epileptoid)" (135, p. 63). *Result:* mainly negative.

19. Humm and Wadsworth, discussing the standardization of the Humm-Wadsworth Temperament Scale, admitted that "approximately 25% of normal subjects and 45% of abnormal subjects manifest a tendency to biased responses in marked degree" (138, p. 181). *Result:* negative.

20. Kelly, Miles and Terman gave their Masculinity-Feminity questionnaire (308) to subjects who were told to try to obtain certain scores on them. They reported that "subjects of both sexes are able to shift their scores enorm-

ously in the masculine or feminine direction as a result of instructions to do so" (151, p. 215). *Result:* negative.

21. Kimber administered the California Personality Test to 400 students of psychology at the University of California, telling them first to answer the questions as they believed that a happy and well-adjusted student would answer them and then to answer them for themselves. He concluded that "most students secured high scores on the first test and low scores on the second" (155, p. 540). *Result:* positive.

22. Kinder gave a questionnaire with 30 person items to 42 college women and had them, anonymously, answer these questions by (a) rating themselves, (b) rating the average college girl, and (c) rating the ideal college girl. He noted that "almost invariably each girl had rated herself higher than she had rated the average college girls. In fact, she often approached what she thought the ideal should be. . . . The intimation that is so often made that people are prone to see themselves through rose-colored glasses and others through inverted opera glasses appears to be borne out. Furthermore, the tendency toward overestimation is apparently a systematic one" (156, pp. 533, 535) *Result:* negative.

23. Knight and Franzen had 110 college juniors rate their interests and rate the interests of the ideal junior. They found that "the correlation between these two orders of interests was .46. There is, then, a tendency for the junior to think he is as he should be." Then they had 60 juniors, on a self-analysis test, rate themselves and rate typical students. They reported that "there is a well-marked tendency for a person to overate himself when he compared himself with others, and even when the introspective judgment is independent of comparison with others, this tendency still persists" (158, pp. 204, 212). *Result:* mainly negative.

24. Landis and Katz administered the Bernreuter Inventory to 224 psychiatric patients and then compared their answers with the known facts of their case histories. They found that only "75% of the answers which are given by neurotic patients are verified by their previous history," and that "contradiction, in the sense of lack of insight or of a different impression which the patient holds in opposition to the impression of his associates, is the most frequent source of a lack of agreement between past history records and test results. This factor operates markedly in those individuals who make a low B1-N score" (172, pp. 349, 355). *Result:* mainly negative.

25. Laslett and Bennett gave the Bernreuter Personality Inventory to a group of students whom they later questioned about their tendencies, consciously or unconsciously, to over-rate themselves. They noted that "numerous students felt that one or more of these factors had decreased the accuracy of their report even though they were in cooperative sympathy with the experiment" (174, p. 461). *Result:* mainly negative.

26. Maller gave children tests involving participation in a group project, under anonymous and non-anonymous conditions, and found that they were significantly more cooperative when they had to write their names on their tests than when they were not required to do so. He then had a group of children rank other children and themselves on a ballot, and found that when they had to put their names on the ballot "they tend to cast a greater number of positive votes, to center the positive votes on a few and to vote less often for themselves" (915, pp. 883–4). *Result:* mainly negative.

27. Manzer compared scores on the Allport Ascendance-Submission Test

when group ratings and individual self-ratings were given. He found that "the distribution of the test scores is nearer that of the self-ratings than it is to that of the group ratings. The inference is drawn that the tendency to make 'flattering' scores, admittedly present in self-ratings, is also present to a considerable extent in the scores made on the Allport Test" (**201**, p. 268). *Result:* mainly negative.

28. Marsh gave a questionnaire to students who were asked to respond to its items both personally and impersonally. He reported that "the most significant single finding of this study is that a few spoken words, at the time a questionnaire is being administered, can make a substantial difference in the results obtained. With the questionnaire here used, results indicate that students are likely to project their own difficulties onto others when asked to respond *impersonally*, and are likely to rationalize when asked to respond *personally*. These results emphasize the necessity for great care in the verbal directions used to supplement any standardized or printed directions, and the need for great caution in the uncritical acceptance of results obtained from the use of questionnaires" (**204**, p. 280). *Result:* mainly negative.

29. Neprash gave the Thurstone Personality Schedule to a group of subjects and re-tested them after intervals of two weeks, four weeks, and eight weeks. He found "a general tendency for scores at the second presentation to indicate a lower degree of maladjustment than those at the first presentation." He also reported that "the responses of the less-well adjusted subjects tended to be more unreliable than those of the better adjusted" (**225**, pp. 241, 243). *Result:* mainly negative.

30. Olson gave the Woodworth-Mathews test to 100 women under anon- and non-anonymous conditions and found that "there is a high probability that more symptoms will be reported in an initial application of the instrument when names are omitted" (**228**, p. 443). *Result:* negative.

31. Robertson and Stromberg gave the Rogers Personality Inventory to college girls after two friends had rated each respondent on the traits it endeavors to diagnose. They reported that "although differences do appear, the over-estimation and the under-estimations cancelled each other out for the group as a whole. This would indicate that over-estimation in the self-rating of college junior and senior women is not a common characteristic" (**253**, p. 127). *Result:* positive.

32. Ruch gave the California Test of Personality to 155 adults under clinical and under employment conditions. He found that the means of the uncorrected scores of those taking the tests under clinical conditions was 132 (55th percentile rank); while the mean of those taking it under employment conditions was 160 (94th percentile rank)" (**253**, p. 540). *Result:* negative.

33. Sailer and Luh gave four self-rating tests to Chinese students, and noted that their results "show uniform and decided tendency toward over-estimation of self as compared with others, giving results comparable to those found with American students" (**269**, p. 249). *Result:* negative.

34. Schott administered the Thurstone test to applicants for professional positions and found that over two-thirds of the subjects made above-average scores. He concluded: "The applicants knew what answers should produce the best impression and accordingly their scores are better measures of sophistication than of the degree of emotional stability. Thus, the alleged weaknesses

of the self-rating type of test are clearly emphasized" (**273**, pp. 238–9). *Result:* negative.

35. Shen gave tests of a self-rating type to 26 subjects and also had them rated by their associates. He found that "our expectations are entirely fulfilled. Of the 26 individuals, six have negative systematic errors and 20 have positive systematic errors. . . . Thus the apparent inaccuracy of self-estimation is largely due to a systematic error of the individual—a systematic tendency over- or under-estimate himself in all traits according to the kind of delusion he has about himself" (**276**, p. 106). *Result:* mainly negative.

36. Simpson gave personality questionnaires to college students and to penitentiary inmates, and reported that "there is a tendency for the penitentiary inmates to rate themselves higher than college students in the possession of generally desirable personality traits" (**279**, pp. 476–7). *Result:* negative.

37. Smith used questionnaires with factual and personal items on a group of subjects who were later re-tested on the same schedules. He observed that "about one-quarter of the items were responded to differently on the two trials, the greatest change being on subjective personal items, the least on factual personal items" (**284**, p. 713). *Result:* probably negative.

38. Spencer gave a personality questionnaire to high school juniors and seniors, and let them fill these out on an anonymous basis. Later he asked if they would have answered the same way if their signatures had been required. He found that only 55.2% of the students reported unequivocally that they would have answered the questions truthfully on a non-anonymous basis. He concluded that "had the instrument been taken under signatures, not only would a large number of the subjects have resented, evaded and falsified some of the items, but such deception would have been greatest among those having the greatest amount of conflict. In short, the purpose of the instrument—the measurement of conflict—would have been invalidated" (**290**, pp. 33, 38). *Result:* mainly negative.

39. Van Wagenen gave the Maller Character Sketches to junior high school students and administered it with several different sets of instructions. He found that when the students were warned that the test results would be put to practical use the obtained differences in scores were "possibly due to a defensive attitude and dissimulation, an effort to respond to questions in such a way as to gain greater approval." He also found that obtained differences when tests were given individually rather than collectively were "possibly due to an attitude of willingness to be frank about oneself and to put aside dissimulation in the presence of the examiner" (**328**,p p. 46–47). *Result:* probably negative.

40. Warner and Gallico interviewed 200 patients who had filled out the Cornell Service Index, and stated that "it was found that less than 2% of that group gave evidence of having answered the questions falsely in order to create an impression of the absence of nervousness" (**331**, pp. 214–215). *Result:* positive.

41. Wasson gave the Humm-Wadsworth to 93 male subjects and reported that, according to its own lie score, "a rather high percentage of the Humm-Wadsworth Scales were answered in an invalid manner" (**334**, p. 362). *Result:* mainly negative.

42. Zapf had public school pupils respond to superstitions on a paper and pencil test and then respond to the same superstitions in actual situations. She found that their responses "were consistent to the extent of 75.49%. Since

the coefficient of correlation between the scores for the two types of response was found to be .79±.03, it would seem to indicate that the paper and pencil testing method is reasonably valid as a measure of superstitiousness, although it is not possible to say that it is *as* good as measurement in actual situations" (**356**, p. 24). *Result:* mainly positive.

In sum: Of 42 experimental attempts to discover whether or not respondents over-rated themselves on self-description instruments, 6 investigators found that they did not, and 36 found that they did.

VALIDITY OF INDIVIDUALLY ADMINISTERED PERSONALITY QUESTIONNAIRES

Most personality questionnaires are designed for administration to groups of subjects rather than to one respondent at a time. A notable exception to this rule has been the Minnesota Multiphasic Personality Inventory (**121**, **122**, **188**, **189**, **190**, **191**), which to date has usually been administered individually. It is questionable whether individual administrations can be directly compared to group testings because, in a sense, they reduce the anonymity, or the pseudo-anonymity, of the testees. A respondent who knows that he alone is being given a personality inventory, and that his score is going to be checked soon afterwards by the administrator of the test (who may also be his teacher or psychologist or personnel adviser) may answer the test items differently than he would have done in an impersonal group administration. Consequently, we have collected in this separate section reports on the validity of personality tests which were individually administered. These include the following:

1. Benton gave the Multiphasic Inventory to 85 neuropsychiatric patients and found that five out of ten schizophrenics gave positive results on its *schizophrenia* scale; five out of nine hysterical patients gave positive results on its *hysteria* scale; 13 out of 16 delinquents gave positive results on its *psychopathic deviate* scale; and nine out of ten homosexuals gave positive results on its *Femininity* scale (**22**, p. 417). *Results:* questionably positive.

2. Benton and Probst administered the Multiphasic Inventory to 70 patients who had been diagnosed by naval psychiatrists. They reported that "in the case of the Psychopathic Deviate, Paranoia and Schizophrenia trends the differences with respect to mean test score between the normal and the abnormal groups can be considered to be statistically significant. . . . On the other hand, there is no substantial amount of agreement with respect to the strength of the Hypochondriasis, Depression, Hysteria, Femininity, and Psychasthenia trends, although in all cases the mean test scores of the psychiatrically rated abnormal groups are higher than those of the normal groups. Especially noteworthy is a complete lack of agreement with respect to the strength of the Hysteria trend" (**23**, p. 76). *Result:* questionably positive.

3. Capwell gave the Multiphasic to a group of delinquent and nondelinquent adolescent girls and observed that "each scale except the Hy, or Hys-

teria scale, shows a clear differentiation between the two groups. The greatest difference appears in the scores for Pd (Psychopathic Deviate)" (**47**, p. 291). *Result:* positive.

4. Gouch used the Multiphasic Inventory with 136 neuropsychiatric army cases and found that the scores of normals basically differed from those having mild psychoneurosis, moderate psychoneurosis, psychopathic personality, severe neurosis, and psychosis. The diagnosis of psychotics was not as accurate, however, as was that of neurotic groups (**98**, p. 37). *Result:* positive.

5. Harris and Christiansen employed the Multiphasic Inventory with 53 patients who had been diagnosed by experienced therapists. They concluded that "the most meaningful way to express the agreement with ratings of experienced therapists is to say that there is perfect agreement in more than half the cases; in another third, there is one-step disagreement, e.g., a Multiphasic rating of *Good* and a clinical rating of *Indifferent;* and in only about 10% of the cases is there complete disagreement" (**117**, p. 276). *Result:* mainly positive.

6. Hathaway and McKinley gave the Multiphasic Inventory to 35 diagnosed depressive cases and to normals, and noted that "although there is some overlap the scale yields scores that differentiate at least 50% of the test cases from normals and even from other psychiatric cases although the latter are, reasonably enough, more depressed than normals" (**122**, p. 81). *Result:* questionably positive.

7. Leverenz used the Multiphasic Inventory for a year at the Fort Snelling Station hospital, and found it to be "of definite value" despite the fact that "the clinical impression was not always corroborated by the scores obtained on the inventory" (**177**, pp. 628–629). *Result:* mainly positive.

8. McKinley and Hathaway gave the Multiphasic Inventory to 25 clinically diagnosed hypochondriasis cases and to a group of normals. They reported that "significant separation of hypochondriacal cases from normals can be demonstrated. . . . A sizeable proportion of the so-called normal population nevertheless overlaps the hypochondriacal group as is often stated in medical and psychiatric opinion" (**188**, p. 267). *Result:* mainly positive.

9. McKinley and Hathaway presented evidence that the Multiphasic Test has valid scales for hysteria, hypomania and psychopathic deviation. They stated that "a scale for Hy for aid in identification of hysterical tendency has been derived. . . . In spite of the small number of criterion and test cases available, a scale for hypomania is presented. . . . A final scale has been developed which will identify half or somewhat more of the cases routinely classed psychopathic personality" (**191**, pp. 162, 167, 173). *Result:* mainly positive.

10. McKinley and Hathaway, using the Multiphasic Schedule on normals and on diagnosed psychasthenics, reported that "through the differential study of persons having psychiatric evidence of psychasthenia, a scale was derived which is internally homogeneous and which differentiates clinic patients from normals in a large percentage of cases" (**190**, p. 624). *Result:* mainly positive.

11. Michael and Buhler gave the Multiphasic Test to 90 neuropsychiatric patients in a public general hospital and found it only 45% successful. They stated that its range of successful diagnosis was limited, and that it was of little help in discriminating psychopaths from psychoneurotics (**215**, pp. 205–211). *Result:* mainly negative.

12. Pai, Sung and Hsu gave the Thurstone Schedule, orally and individu-

ally, to 127 normals, 224 neurological patients, 62 applicants for non-professional work, 101 mental patients, and 103 drug addicts. They report that "the score of the neurological patients was significantly higher than that of the medical patients [normals], applicants, and the addicts. The scores of the mental patients and those of the addicts and medical patients [normals] were significantly higher than those of the applicants" (**231**, p. 71). *Result:* mainly positive.

13. Schmidt gave the Multiphasic Inventory to normal and abnormal Army Air Force personnel and found that it did "distinguish graphically and with statistical significance between normal soldiers and those diagnosed as constitutional psychopaths; mild or severe neurosis; and psychosis" (**272**, pp. 130–131). *Result:* positive.

14. Smith gave a test of neurotic tendency to manic depressive and dementia praecox patients, but had it administered individually by the patients' nurses. She found the two groups to be clearly differentiated, with the dementia praecox patients having significantly higher neurotic scores (**283**, p. 117). *Result:* positive.

15. Van Vorst used the Multiphasic Test with a group of psychopathic delinquents, and reported that the obtained results did not appear to support the claims of the test "to an extent which would justify defining any characteristic response pattern for the psychopathic personality" (**327**, p. 583). *Result:* mainly negative.

In sum: Of 15 studies where personality tests of the conventional type were given individually, instead of in group administrations, the validity results were positive in ten cases, questionably positive in three, and negative in two.

DISCUSSION

If we consider the direct validation experiments of group-administered personality questionnaires—that is to say, investigations where test results were evaluated against outside clinical criteria such as behavior problem diagnosis, delinquency, psychiatric and psychological diagnosis, and ratings by friends and associates—we find that out of 162 reported studies, 65 showed positive, 26 questionably positive, and 71 negative results. If we add to these evaluations reports of experiments where the validity of personality tests was indirectly investigated—that is, by intercorrelating them with similar tests or estimating respondents' over-rating tendencies—we have 97 additional studies, 15 of which give positive, 18 questionably positive, and 64 negative indications of validity. This makes a grand total of 259 investigations (some of which, however, overlap): 80 positive or mainly positive; 44 questionably positive; and 135 negative or mainly negative. Obviously, this is not a very good record for the validity of paper and pencil personality questionnaires.

In those experiments where personality tests of the conventional type were given individually, instead of to groups of respondents, the record of ten positive, three questionably positive, and two negative results is patently far better than the validity record of group-administered tests. There may be several explanations for this advantage:

1. Individual administration may bring about, at least in part, the same kind of rapport factors which are so important in case study interviews.

2. Most individual administrations have been done with one test, the Minnesota Multiphasic, which was standardized on a decidedly clinical and objective, rather than the more usual subjective, basis and which, in consequence, may possibly be a superior questionnaire.

3. The majority of Multiphasic validity studies have either been done on groups used to standardize the test, which have been institutionalized populations which may be more sophisticated and more honest than other abnormal groups; or else they have been done with military personnel, who may have every incentive to answer personality questionnaires honestly.

Whether any of these three hypotheses actually accounts for the seeming superiority of individually-administered over group-administered personality tests is a question that warrants further investigation.

Returning to the group-administered tests whose validation has been mainly reviewed in this paper, we may break down the figures on direct validity studies, to discover whether any single personality questionnaire on which ten or more experiments have been reported has proved to possess high validity. Doing this, we obtain Table II. A glance at this table will show that the most widely used tests all have highly dubious validity records.

It is true that, upon more detailed breakdown, some of the tests obtain better validity records in one field—*e.g.*, psychiatric or psychological diagnosis—than they do in other areas—*e.g.*, behavior problem diagnosis or diagnosis by ratings. But in no case does it seem that any one of the most popular personality questionnaires has been shown to have indisputably high validity in any particular area. However, when the existence of more validation studies makes feasible finer breakdowns of the data, it will be interesting to see whether one of the indications of this study—to wit, that specific personality questionnaires are more valid for some purposes than for others—will be borne out.

It may be asked whether the picture of personality questionnaire validity given by literature reports is actually a fair one. Have experimenters been biased against these tests, and consequently leaned over backwards to obtain negative results in validating them? On the contrary, it must be pointed out that the indications of questionnaire validity summarized in this article are probably higher than they should be.

Many of the studies reporting positive validations are suspect on several counts:

1. They were done by the authors of the tests in question, who frequently went to every extreme, particularly in their interpretations of statistical significance, to obtain favorable results.

2. The groups of subjects employed for validating purposes were often unusually test-sophisticated or biased ones, and could be expected to answer more honestly than normal subjects. Thus, they on many occasions consisted of (a) institutionalized patients who were used to taking personality tests or had nothing to lose by taking them; or (b) individuals who voluntarily came for psychological aid, and who would be expected to give unusually honest answers; or (c) army personnel who were under decided compulsion (as Hunt and Stevenson (141) incisively pointed out) to answer questionnaires honestly.

TABLE II

VALIDITY STUDIES OF PERSONALITY QUESTIONNAIRES

Test Employed	*Number of Times Employed*	*Number of Positive Validations*	*Number of Questionably Positive Validations*	*Number of Negative Validations*
Bell Adjustment Inventory	12	1	0	11
Bernreuter Personality Inventory	29	9	6	14
Thurstone Personality Schedule	10	4	1	5
Woodworth Personal Data Sheet	29	11	4	14
Other Personality Tests	82	40	15	27
Totals	162	65	26	71

3. The majority of "significant differences" reported in the positive validations reviewed in this paper were only differences between the

mean scores of *groups* taking the tests under consideration; and the correlations denoted only *group* relationships and were not coefficients of dependability or of forecasting efficiency. Consequently, most of the tests investigated, even when they were shown to be in some degree valid, were *not* shown to be so for purposes of *individual* diagnosis—which is, after all, the one main purpose for which personality questionnaires are presumably constructed.

We may conclude, therefore, that judging from the validity studies on group-administered personality questionnaires thus far reported in the literature, there is at best one chance in two that these tests will validly discriminate between *groups* of adjusted and maladjusted individuals, and there is very little indication that they can be safely used to diagnose *individual* cases or to give valid estimations of the personality traits of *specific* respondents. The older, more conventional, and more widely used forms of these tests seem to be, for practical diagnostic purposes, hardly worth the paper on which they are printed. Among the newer questionnaires, the Minnesota Multiphasic schedule appears to be the most promising one—perhaps because it gets away from group administration which has hitherto been almost synonymous with personality test-giving. More research in this direction is well warranted at the present time.

SUMMARY

A review of the literature on validity studies of personality questionnaires has been made and the following points brought out:

1. Of 9 attempts to validate personality tests on behavior problem children, 2 gave positive, 1 questionably positive, and 6 negative results.
2. Of 34 attempts to validate them with delinquents, 15 gave positive, 6 questionably positive, and 13 negative results.
3. Of 75 attempts to evaluate them against psychiatric and psychological diagnoses, 36 brought forth positive, 9 questionably positive, and 30 negative results.
4. Of 44 attempts to check them against ratings by friends, associates, or teachers of respondents, 12 gave positive, 10 questionably positive, and 22 negative results.
5. Of 55 attempts to correlate similar personality tests with each other, 9 gave positive, 18 questionably positive, and 28 negative indications of validity.
6. Of 42 investigations of over-rating on questionnaires, 6 showed that respondent over-estimation did not, and 36 showed that it did, take place.
7. Of 15 validity studies where personality questionnaires were ad-

ministered individually, instead of in group administrations, 10 gave positive, 3 questionably positive, and 2 negative results.

8. It is concluded that group-administered paper and pencil personality questionnaires are of dubious value in distinguishing between groups of adjusted and maladjusted individuals, and that they are of much less value in the diagnosis of individual adjustment or personality traits.

BIBLIOGRAPHY

1. Adams, C. R. A new measure of personality. *J. appl. Psychol.*, 1941, **25**, 141–151.
2. Adams, C. R. *Manual of directions for using and interpreting the Personal Audit, forms LL and SS.* Chicago: Science Research Associates, 1945.
3. Adams, C. R. & Lepley, W. M. *The Personal Audit.* State College, Penna.
4. Alexander, F. Evaluation of statistical and analytical methods in psychiatry and psychology. *Amer. J. Orthopsychiat.*, 1934, **4**, 433–438.
5. Allen, Eleanor A. Temperamental Tests. *Brit. J. med. Psychol.*, 1927, **7**, 394–446.
6. Allport, G. W. Concepts of trait and personality. *Psychol. Bull.*, 1927, **24**, 284–293.
7. Allport, G. W. A test for ascendance-submission. *J. abnorm. soc. Psychol.*, 1928, **23**, 118–136.
8. Allport, G. W. The neurotic personality and traits of self-expression. *J. soc. Psychol.*, 1930, **1**, 524–527.
9. Allport, G. W. *Personality: a psychological interpretation.* New York: Henry Holt, 1937.
10. Allport, G. W. & Vernon, P. E. The field of personality. *Psychol. Bull.*, 1930, **27**, 677–730.
11. Anderson, Rose G. Some technological aspects of counseling adult women. *J. appl. Psychol.*, 1938, **22**, 455–469.
12. Ascher, E. J. & Haven, S. E. The reactions of state correctional school and public school boys to the questions of an emotional inventory. *J. juv. res.*, 1930, **14**, 96–106.
13. Babcock, Marjorie E. *A comparison of delinquent and non-delinquent boys by objective measures of personality.* Honolulu: Author, 1932.
14. Bain, R. Stability in questionnaire response. *Amer. J. Sociol.*, 1931, **37**, 445–453.
15. Bartlett, E. R. & Haas, D. B. Personality factors in delinquency. *Sch. & Soc.*, 1936, **43**, 563–656.
16. Beckman, R. O. Ascendance-submission test revised. *Person. J.*, 1933, **11**, 387–393.
17. Bell, H. M. *The Adjustment Inventory.* Stanford: Stanford University Press, 1938.
18. Bell, H. M. Personal counseling: using the adjustment inventory in counseling. *Report of the 13th Annual Meeting of the American College Personnel Assoc.*, 1936, 31–34.
19. Bennett, Elisabeth. Some tests for the discrimination of neurotic from normal subjects. *Brit. J. med. Psychol.*, 1945, **20**, 271–277.
20. Bennett, G. K. & Gordon, H. P. Personality test scores and success in the field of nursing. *J. appl. Psychol.*, 1944, **28**, 267–278.
21. Benton, A. L. The interpretation of questionnaire items in a personality schedule. *Arch. Psychol. N. Y.* No. 190, 1935.
22. Benton, A. L. The Minnesota

Multiphasic Inventory in clinical practice. *J. nerv. Mental Dis.*, 1945, **102**, 416–420.

23. Benton, A. L. & Probst, Kathryn. A comparison of psychiatric ratings with Minnesota Multiphasic Personality Inventory Scores. *J. abnorm. soc. Psychol.*, 1946, **41**, 75–78.
24. Bernreuter, R. G. The theory and construction of the personality inventory. *J. soc. Psychol.*, 1933, **4**, 383–405.
25. Bernreuter, R. G. Validity of the presonality inventory. *Person. J.*, 1933, **11**, 383–386.
26. Bernreuter, R. G. The imbrication of tests of introversion-extroversion and neurotic tendency. *J. soc. Psychol.*, 1934, **5**, 184–199.
27. Bernreuter, R. G. The measurement of self-sufficiency. *J. abnorm. soc. Psychol.*, 1934, **28**, 291–300.
28. Bernreuter, R. G. Chance and personality. *J. educ. Psychol.*, 1935, **26**, 279–283.
29. Bernreuter, R. G. *The personality inventory.* Stanford: Stanford University Press, 1935.
30. Bingham, W. V. Reliability, validity and dependability. *J. appl. Psychol.*, 1932, **16**, 116–122.
31. Bonney, M. E. The validity of certain techniques of gathering psychological data, with special reference to personality questionnaires. *J. soc. Psychol.*, 1941, **13**, 103–122.
32. Boynton, P. L. & Walsworth, B. M. Emotionality test scores of delinquent and non-delinquent girls. *J. cbnorm. soc. Psychol.*, 1943, **38**, 87–92.
33. Bridges, J. W. A study of a group of delinquent girls. *J. genet. Psychol.*, 1927, **34**, 187–204.
34. Bridges, J. W. & Bridges, K. M. B. A psychological study of juvenile delinquency by group methods. *Genet. Psychol. Monogr.*, 1926, **1**, 411–506.
35. Broom, M. E. A critical study of a test of extroversion-introversion traits. *J. juv. Res.*, 1929, **13**, 104–123.
36. Broom, M. E. A study of a test of Ascendance-Submission. *J. appl. Psychol.*, 1930, **14**, 405–413.
37. Brown, F. *Personality inventory for children.* New York: Psychological Corp., 1935.
38. Brown, F. An experimental study of the validity and reliability of the Brown Personality Inventory for children. *J. Psychol.*, 1944, **17**, 75–89.
39. Burks, Barbara S. Personality scores in relation to measurement. *J. soc. Psychol.*, 1936, **7**, 140–149.
40. Buros, O. K. *Educational, psychological and personality tests of 1933, 1934, and 1935.* New Brunswick N. J.: School of Education, Rutgers University, 1936.
41. Buros, O. K. *Educational, psychological and personality tests of 1936.* New Brunswick, N. J.: School of Education, Rutgers University, 1937.
42. Buros, O. K. *The 1938 mental measurements yearbook of the school of Education.* Rutgers University. New Brunswick, N. J.: Rutgers University, 1938.
43. Buros, O. K. *The 1944 mental measurements yearbook.* Highland Park, N. Y.: Mental Measurements Yearbook, 1941.
44. Cady, V. M. The estimation of juvenile incorrigibility. *J. Delinqu. Monogr.*, 1923, No. 2.
45. Campbell, Katherine J. The application of extroversion-introversion tests to the insane. *J. abnorm. soc. Psychol.*, 1929, **23**, 479–481.
46. Capwell, Dora F. Personality patterns of adolescent girls: I. Girls who show improvement in IQ. *J. appl. Psychol.*, 1945, **29**, 212–228.
47. Capwell, Dora F. Personality patterns of adolescent girls: II. De-

linquents and non-delinquents. *J. appl. Psychol.*, 1945, **29**, 289–297.

48. CASSELBERRY, W. S. Analysis and prediction of delinquency. *J. juv. Res.*, 1932, **16**, 1–31.

49. CASSELBERRY, W. S. Personality measurement from the standpoint of a counseling psycholgoist. *Psychol. Bull.*, 1934, **31**, 583.

50. CATTELL, R. B. *A guide to mental testing.* London: University of London Press, 1936.

51. CAVAN, RUTH S. The Murray Psychoneurotic Inventory and the White House Conference Inventory. *J. juv. Res.*, 1934, **18**, 23–27.

52. CLARK, W. A. & SMITH, L. F. Further evidence on the validity of personality inventories. *J. educ. Psychol.*, 1942, **33**, 81–91.

53. COGAN, LUCY C., CONKLIN, AGNES M., & HOLLINGWORTH, H. L. An experimental study of self-analysis, estimates of associates, and the results of tests. *Sch. & Soc.*, 1915, **2**, 171–179.

54. CONKLIN, E. S. The definition of introversion-extroversion and allied concepts. *J. abnorm. soc. Psychol.*, 1923, **17**, 367–382.

55. CONKLIN, E. S. The determination of normal extravert-introvert differences. *J. genet. Psychol.*, 1927, **34**, 28–37.

56. CONWAY, C. B. A new scoring apparatus for the Bernreuter Personality Inventory. *J. appl. Psychol.*, 1936, **20**, 264–265.

57. COOK, P. H. The clinical testing of personality. *Aust. J. Psychol. Phil.*, 1939, **7**, 151–157.

58. COPELAND, H. A. Studies in the reliability of personnel records. *J. appl. Psychol.*, 1938, **22**, 247–251.

59. COURTHIAL, ANDREE. Emotional differences of delinquent and non-delinquent girls of normal intelligence. *Arch. Psychol., N. Y.*, No. 133, 1931.

60 CROOK, M. N. Differential responses to personality test items. *Psychol. Rec.*, 1937, **1**, 187–194.

61. CUSHING, HAZEL M. & RUCH, G. M. An investigation of character traits in delinquent girls. *J. appl. Psychol.*, 1927, **11**, 1–7.

62. DANIEL, R. P. *A psychological study of delinquent and non-delinquent negro boys.* New York: Bureau of Publications, Teachers College, Columbia University, 1932.

63. DARLEY, J. G. Tested maladjustment related to clinically diagnosed maladjustment. *J. appl. Psychol.*, 1937, **21**, 632–642.

64. DARLEY, J. G. & ANDERSON, G. V. Applications of personality and character measurement. *Rev. educ. Res.*, 1944, **14**, 67–80.

65. DARLEY, J. G. & INGLE, D. J. An analysis of the Bernreuter Personality Inventory in occupational guidance. *Bull. Employment Stabilization Res. Inst., Univ. Minn.*, 1934, **3**, 32–41.

66. DARLEY, J. G. & MCNAMARA, W. J. Factor analysis in the establishment of new personality tests. *J. educ. Psychol.*, 1940, **31**, 321.

67, DE ANGELIS, L. *A study of the Bernreuter test with psychiatric cases.* Unpublished M.A. Thesis, Columbia University, 1932.

68. DRAKE, MARGARET J., ROSLOW, S., & BENNETT, G. K. The relationship of self-rating and classmate rating on personality traits. *J. exp. Educ.*, 1939, **7**, 210–213.

69. DUDYCHA, G. J. Self-estimates and dependability. *J. soc. Psychol.*, 1940, **12**, 39–53.

70. DUREA, M. A. & BILSKY, H. B. An exploratory study of personality characteristics in schizophrenia and manic-depressive psychoses. *J. genet. Psychol.*, 1943, **28**, 81–98.

71. EISENBERG, P. Individual interpretation of psychoneurotic inventory items. *J. gen. Psychol.*, 1941, **25**, 19–40.

72. Eysenck, H. J. The effect of incentives on neurotics, and the variability of neurotics as compared with normals. *Brit. J. med. Psychol.*, 1944, **20**, 100–103.

73. Eysenck, H. J. A comparative study of four screening tests for neurotics. *Psychol. Bull.*, 1945, **42**, 659–662.

74. Farnsworth, P. R. A study of Bernreuter profiles. *Psychol. Bull.*, 1933, **30**, 600–601.

75. Feder, D. D. & Baer, L. O. A comparison of test records and clinical evaluations of personality and adjustment. *J. educ. Psychol.*, 1941, **32**, 133–144.

76. Feder, D. D. & Mallet, D. R. Validity of certain measures of personality adjustment. *J. Amer. Ass. colleg. Registr.*, 1937, **13**, 5–15.

77. Fisher, V. E. & Watson, R. I. A new approach to the old problem of emotional stability. *Psychol. Bull.*, 1939, **36**, 563–564.

78. Fisher, V. E. & Watson, R. I. An inventory of affective tolerance. *J. Psychol.*, 1941, **12**, 149–157.

79. Fisher, W. & Hayes, S. P. Maladjustment in college, predicted by Bernreuter Inventory scores and family position. *J. appl. Psychol.*, 1941, **25**, 86–96.

80. Flanagan, J. C. *Factor analysis in the study of personality.* Stanford: Stanford University Press, 1935.

81. Flanagan, J. C. Technical aspects of multi-trait tests. A reply to Dr. Lorge. *J. educ. Psychol.*, 1935, **26**, 641–653.

82. Flemming, E. G. The predictive value of certain tests of emotional stability as applied to college freshmen. *Arch. Psychol. N. Y.*, No. 96, 1928.

83. Flemming, E. G. Emotion, intelligence and introversion. *J. gen. Psychol.*, 1932, **7**, 201–205.

84. Flemming, E. G. & Flemming, Cecile W. The validity of the Mathews' Revision of the Woodworth Personal Data Questionnaire. *J. abnorm. soc. Psychol.*, 1929, **23**, 500–506.

85. Frank, B. Stability of questionnaire response. *J. abnorm. soc. Psychol.*, 1936, **30**, 320–324.

86. Frank, L. K. Projective methods for the study of personality. *J. Psychol.*, 1939, **8**, 389–413.

87. Franz, S. I. *Handbook of Mental Examination Methods* (2nd Ed. revised). New York: Macmillan, 1919.

88. Freeman, F. N. Tests of personality traits. *Sch. Rev.*, 1936, **33**, 95–106.

89. Garrett, H. E. *Statistics in Psychology and Education* (2nd Ed.). New York: Longmans, Green, 1940.

90. Garrett, H. E. & Schneck, M. R. A study of the discriminative value of the Woodworth Personal Data Sheet. *J. genet. Psychol.*, 1928, **1**, 459–471.

91. Garrett, H. E. & Schneck, M. R. *Psychological tests, methods and results.* New York: Harper, 1933.

92. Gibb, C. A. Personality traits by factorial analysis. *Aust. J. Psychol. Phil.*, 1942, **20**, 1–15.

93. Gilliland, A. R. Problems of personality. *J. abnorm. soc. Psychol.*, 1928, **23**, 369–378.

94. Gilliland, A. R. What do introversion-extroversion tests measure? *J. abnorm. soc. Psychol.*, 1934, **28**, 407–412.

95. Gilliland, A. R. & Morgan, J. J. B. An objective measure of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1931, **26**, 296–303.

96. Gordon, H. C. & Davidoff, P. Honesty of pupils in answering adjustment questionnaires. *Sch. & Soc.*, 1943, **57**, 54–56.

97. Gorham, D. R. & Brotemarkle, R. A. Challenging three standardized emotional tests for validity and employability. *J. appl. Psychol.*, 1929, **13**, 554–581.

98. Gouch, H. G. Diagnostic patterns on the Minnesota Multiphasic Per-

sonality Inventory. *J. clin. Psychol.*, 1946, **2**, 23–37.

99. Greene, J. E. & Staton, T. F. Predictive value of various tests of emotionality and adjustment in a guidance program for prospective teachers. *J. educ. Res.*, 1939, **32**, 653–659.

100. Guilford, J. P. An examination of a typical test of introversion-extroversions by means of the method of similar reactions. *J. soc. Psychol.*, 1933, **4**, 430–443.

101. Guilford, J. P. Introversion-extroversion. *Psychol. Bull.*, 1934, **31**, 331–354.

102. Guilford, J. P. A simple scoring weight for test items and its reliability. *Psychometrika*, 1941, **6**, 367–374.

103. Guilford, J. P. & Braly, K. W. Extroversion and introversion. *Psychol. Bull.*, 1930, **27**, 96–107.

104. Guilford, J. P. & Guilford, Ruth B. An analysis of the factors in a typical test of introversion-extroversion. *J. abnorm. soc. Psychol.*, 1934, **28**, 377–399.

105. Guilford, J. P. & Guilford, Ruth B. Personality factors, S, E, & M, and their measurement. *J. Psychol.*, 1936, **2**, 109–127.

106. Guilford, J. P. & Guilford, Ruth B. Personality factors, D, R, T, and A. *J. abnorm. soc. Psychol.*, 1939, **34**, 21–36.

107. Guilford, J. P. & Guilford, Ruth B. Personality factors N and GD. *J. abnorm. soc. Psychol.*, 1939, **34**, 239–248.

108. Guilford, J. P. & Martin, H. *An Inventory of Factors S T D C R.* Beverly Hills, California: Sheridan Supply Co., 1940.

109. Guilford, J. P. & Martin, H. Age differences and sex differences in some introvertive and emotional traits. *J. gen. Psychol.*, 1944, **31**, 219–229.

110. Guthrie, E. R. Measuring introversion and extroversion. *J. abnorm. soc. Psychol.*, 1927, **22**, 82–88.

111. Hampton, P. Language difficulties of the Bernreuter Personality Inventory. *J. educ. Psychol.*, 1941, **32**, 471–473.

112. Hanks, L. M. Prediction from case material to personality test data. *Arch. Psychol., N.Y.*, No. 207, 1936.

113. Hanna, J. V. Clinical procedure as a method of validating a measure of psychoneurotic tendency. *J. abnorm. soc. Psychol.*, 1934, **28**, 435–445.

114. Harriman, P. L. A follow-up study of the Woodworth-House Mental Hygiene Inventory. *Amer. J. Orthopsychiat.*, 1938, **8**, 255–259.

115. Harris, D. B. The relation of Maller Case Inventory scores to institutional adjustment of delinquent boys. *J. educ. Psychol.*, 1941, **32**, 550–553.

116. Harris, D. B. & Dobelstein, D. A. A study of the Maller and Boynton Personality Inventories. *J. educ. Psychol.*, 1938, **29**, 279–286.

117. Harris, R. E. & Christiansen, Carole. Prediction of response to brief psychotherapy. *J. Psychol.*, 1946, **21**, 269–284.

118. Hartmann, G. W. Measuring teaching efficiency among college instructors. *Arch. Psychol., N.Y.*, No. 154, 1933.

119. Harvey, O. L. Concerning the Thurstone Personality Schedule. *J. soc. Psychol.*, 1932, **3**, 240–251.

120. Hathaway, S. R. The personality inventory as an aid in the diagnosis of psychopathic inferiors. *J. consult. Psychol.*, 1939, **3**, 112–117.

121. Hathaway, S. R. & McKinley, J. C. A Multiphasic Personality Schedule (Minnesota): I. Construction of the schedule. *J. Psychol.*, 1940, **10**, 249–254.

122. Hathaway, S. R. & McKinley, J. C. A Multiphasic Personality Schedule (Minnesota): III. The

measurement of symptomatic depression. *J. Psychol.*, 1942, **14**, 73–84.

123. Heidbreder, Edna. Measuring introversion and extroversion. *J. abnorm. soc. Psychol.*, 1926, **21**, 120–134.
124. Heidbreder, Edna. Introversion and extroversion in men and women. *J. abnorm. soc. Psychol.*, 1927, **22**, 52–61.
125. Heidbreder, Edna. Self-ratings and preferences. *J. abnorm. soc. Psychol.*, 1930, **25**, 62–74.
126. Hildreth, Gertrude H. *A bibliography of mental tests and rating scales.* New York: Psychological Corp., 1939.
127. Hoitsma, R. K. The reliability and relationships of the Colgate Mental Hygiene Test. *J. appl. Psychol.*, 1925, **9**, 293–303.
128. Hollingworth, H. L. *The psychology of functional neuroses.* New York: Appleton, 1920.
129. Horsch, A. C. & Davis, R. A. Mental hygiene and personality tests. *Amer. J. Sociol.*, 1935, **40**, 646–650.
130. Horsch, A. C. & Davis, R. A. Personality traits of juvenile delinquents and adult criminals. *J. soc. Psychol.*, 1938, **9**, 57–65.
131. House, S. D. A mental hygiene inventory. *Arch. Psychol.*, N. Y., 1927, No. 88.
132. Houtchens, H. M. Temperament in adolescent groups. *Univ. Ia. Stud. Child Welf.*, 1938, **15**, 7–68.
133. Humm, D. G. Personality and adjustment. *J. Psychol.*, 1943, **13**, 109–134.
134. Humm, D. G. Discussions of Dorcus' study of the Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1944, **28**, 527–529.
135. Humm, D. G. & Humm, Kathryn A. Validity of the Humm-Wadsworth Temperament Scale, with a consideration of the effects of subjects' response-bias. *J. Psychol.*, 1944, **18**, 55–64.
136. Humm, D. G., Storment, R. C., & Iarns, M. E. Combination scores for the Humm-Wadsworth Temperament Scale. *J. Psychol.*, 1939, **7**, 227–253.
137. Humm, D. G. & Wadsworth, G. A diagnostic inventory of temperament, preliminary report. *Psychol. Bull.*, 1933, **30**, 602.
138. Humm, D. G. & Wadsworth, G. W. The Humm-Wadsworth Temperament Scale. *Amer. J. Psychiat.*, 1935, **92**, 163–200.
139. Humm, D. G. & Wadsworth, G. W. Using the Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1941, **25**, 654–659.
140. Hunt, Thelma. The measurement of social intelligence. *J. appl. Psychol.*, 1928, **12**, 317–334.
141. Hunt, W. & Stevenson, Iris. Psychological testing in military clinical psychology: II. Personality testing. *Psychol. Rev.*, 1946, **53**, 107–115.
142. Hunt, W., Wittson, C. L., & Harris, H. I. The screen test in military selection. *Psychol. Rev.*, 1944, **51**, 37–46.
143. Ingles, D. J. A test of mental instability. *J. appl. Psychol.*, 1934, **18**, 252–266.
144. Jackson, J. A mutual validation of personality traits. *J. soc. Psychol.*, 1945, **22**, 195–202.
145. Jackson, J. The relative effectiveness of paper-pencil test, interview, and ratings as techniques for personality evaluation. *J. soc. Psychol.*, 1946, **23**, 35–54.
146. Jackson, R. B. & Ferguson, G. A. *Studies on the reliability of tests.* Toronto: University of Toronto, 1941.
147. Jarvie, L. L. & Johns, A. A. Does the Bernreuter Personality Inventory contribute to counseling. *Ohio State Educ. Res. Bull.*, 1938, **17**, 7–9.
148. Jasper, H. H. The measurement of

depression-elated and its relation to a measure of extraversion-introversion. *J. abnorm. soc. Psychol.*, 1930, **25**, 307–318.

149. Johnson, Hope. Objective tests, including the Bernreuter Personality Inventory, as constructive elements in a counseling technique. *Amer. J. Orthopsychiat.*, 1936, **6**, 431–436.

150. Johnson, W. B. The effect of mood on personality traits as measured by Bernreuter. *J. soc. Psychol.*, 1934, **5**, 515–522.

151. Kelly, E. L., Miles, Catherine C., & Terman, L. M. Ability to influence one's score on a typical pencil and paper test of personality. *Character & Pers.*, 1935–1936, **4**, 206–215.

152. Kempfer, H. Simplifying the scoring techniques of the Bernreuter Personality Inventory. *J. appl. Psychol.*, 1944, **28**, 412–413.

153. Keys, N. & Guilford, Margaret S. The value of certain adjustment inventories in predicting problem behavior. *J. educ. Psychol.*, 1937, **28** 641–655.

154. Keys, N. & Whiteside, G. H. The relation of nervous-emotional stability to educational achievement. *J. educ. Psychol.*, 1930, **21**, 429–441.

155. Kimber, M. Insight of college students into the items of a personality test. *Psychol. Bull.*, 1945, **42**, 540.

156. Kinder, J. S. Through our own looking-glass. *Sch. & Soc.*, 1925, **22**, 533–536.

157. Kirkpatrick, C. A critical note on the statistical study of personality reactions. *J. abnorm. soc. Psychol.*, 1933, **28**, 168–171.

158. Knight, F. B. & Franzen, R. H. Pitfalls in rating schemes. *J. educ. Psychol.*, 1922, **13**, 204–213.

159. Koos, L. V. *The questionnaire in education.* New York: Macmillan, 1923.

160. Kornhauser, A. Replies of psychologists to a short questionnaire on mental test developments, personality inventories, and the Rorschach test. *Educ. Psychol. Msmt.*, 1945, **5**, 3–15.

161. Krait, M. H. Personality-testing in the light of the situational approach. *Amer. J. Psychiat.*, 1931, **10**, 839–854.

162. Kruger, Barbara L. A statistical analysis of the Humm-Wadsworth Temperament Scale. *J. appl. Psychol.*, 1938, **22**, 641–652.

163. Kuznets, G. An analysis of Bernreuter's Personality Inventory. *Psychol. Bull.*, 1934, **31**, 585.

164. Kuznets, G. M. An analysis of a group of most differentiating items of the Thurstone Personality Schedule. *Psychol. Bull.*, 1938, **35**, 525.

165. Laird, D. A. Detecting abnormal behavior. *J. abnorm. soc. Psychol.*, 1925, **20**, 128–141.

166. Laird, D. A. A mental hygiene and vocational test. *J. educ. Psychol.*, 1925, **16**, 419–422.

167. Laird, D. A. *Personal Inventory.* Hamilton, N. Y.: Hamilton Republican, 1925.

168. Landis, C. An attempt to measure emotional traits in juvenile delinquency. In *Studies in the Dynamics of Behavior*, edited by Karl S. Lashley. Chicago: University of Chicago Press, 1932.

169. Landis, C. Questionnaires and the study of personality. *J. nerv. ment. Dis.*, 1936, **83**, 125–134.

170. Landis, C., Ferrall, Sarah, & Page, J. Fear, anger and disease: their intercorrelations in normal and abnormal people. *Amer. J. Psychol.*, 1936, **48**, 585–597.

171. Landis, C., Gullette, Ruth, & Jacobsen, C. Criteria of emotionality. *J. genet. Psychol.*, 1925, **32**, 209–234.

172. Landis, C. & Katz, S. E. The valid-

ity of certain questionnaires which purport to measure neurotic tendencies. *J. appl. Psychol.*, 1934. **18**, 343–356.

173. LANDIS, C., ZUBIN, J., & KATZ, S. E. Empirical evaluation of three personality adjustment inventories. *J. educ. Psychol.*, 1935, **26**, 321–330.

174. LASLETT, H. R. & BENNETT, ELISABETH. A comparison of scores on two measures of personality. *J. abnorm. soc. Psychol.*, 1934, **28**, 459–461.

175. LAYMAN, EMMA M. An item analysis of the adjustment questionnaire. *J. Psychol.*, 1940, **10**, 87–106.

176. LENTZ, T. F. Reliability of opinionaire technique studied intensively by the retest method. *J. soc. Psychol.*, 1934, **5**, 338–364.

177. LEVERENZ, C. W. Minnesota Multiphasic Personality Inventory. *War Med.*, 1943, **4**, 618–629.

178. LINK, H. C. A test of four personality traits of adolescents. *J. appl. Psychol.*, 1936, **20**, 527–534.

179. LINK, H. C. *Inventory of Activities and Interests (Personality Quotient Test)*, 1938 Revision. New York: Psychological Corp., 1938.

180. LOOFBOUROW, G. C. & KEYS, N. *Personal Index.* Minneapolis: Educational Test Bureau, 1937.

181. LORGE, I. Personality traits by fiat: I. The analysis of the total trait scores and keys of the Bernreuter Personality Inventory. *J. educ. Psychol.*, 1935, **26**, 273–278.

182. LORGE, I. Personality traits by fiat: II. A correction. *J. educ. Psychol.*, 1935, **26**, 652–654.

183. LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545–546.

184. LORGE, I., BERNHOLZ, ELVA, & SELLS, S. F. Personality traits by fiat: III. The consistency of the Bernreuter Personality Inventory by the Bernreuter and Flanagan keys. *J. educ. Psychol.*, 1935, **26**, 427–454.

185. MCCARTNEY, J. L. & PAPURT, M. J. Preliminary report on an objective personal data questionnaire. *J. abnorm. soc. Psychol.*, 1935, **31**, 95–106.

186. MCFARLAND, R. A. & SEITZ, C. P. *P-S Experience Blank (Psychosomatic Inventory).* New York: Psychological Corp., 1938.

187. MCFARLAND, R. A. & SEITZ, C. P. A Psycho-Somatic Inventory. *J. appl. Psychol.*, 1938, **22**, 327–339.

188. MCKINLEY, J. C. & HATHAWAY, S. R. A Multiphasic Personality Schedule (Minnesota): II. A differential study of hypochrondriasis. *J. Psychol.*, 1940, **10**, 255–268.

189. MCKINLEY, J. C. & HATHAWAY, S. R. A Multiphasic Personality Schedule (Minnesota): IV. Psychasthenia. *J. appl. Psychol.*, 1942, **26**, 614–624.

190. MCKINLEY, J. C. & HATHAWAY, S. R. The identification and measurement of the psychoneuroses in medical practice. *J. Amer. med. Ass.*, 1943, **122**, 161–167.

191. MCKINLEY, J. C. & HATHAWAY, S. R. The Minnesota Multiphasic Personality Inventory. V. Hysteria, hypomanic and psychopathic deviate. *J. appl. Psychol.*, 1944, **28**, 153–174.

192. MCKINNEY, F. Concomitants of adjustment and maladjustment in college students. *J. abnorm. soc. Psychol.*, 1937, **31**, 435–457.

193. MCNEMAR, OLGA W. & LANDIS, C. Childhood disease and emotional maturity in the psychopathic woman. *J. abnorm. soc. Psychol.*, 1935, **30**, 314–319.

194. MCQUITTY, L. L. Conditions affecting the validity of personality inventories: I. *J. soc. Psychol.*, 1942, **15**, 33–39.

195. MALLER, J. B. The effect of signing one's name. *Sch. & Soc.*, 1930, **31**, 882–884.

196. Maller, J. B. Character and personality tests. *Psychol. Bull.*, 1934, **31**, 501–520.

197. Maller, J. B. Character and personality tests. *Psychol. Bull.*, 1935, **32**, 500–523.

198. Maller, J. B. *Character Sketches.* New York: Bureau of Publications, Teachers College, Columbia University, 1936.

199. Maller, J. B. *Character and personality tests.* New York: Bureau of Publications, Teachers College, Columbia University, 1937.

200. Maller, J. B. Personality tests. In Hunt, J. Mc. V. (Ed.), *Personality and the behavior disorders.* New York: Ronald Press, 1944.

201. Manzer, C. W. The effects of self-interest on scores made on the Allport Test for measuring Ascendance-Submission. *Psychol. Clin.*, 1932, **21**, 268–272.

202. Marsh, C. J. The diagnostic value of the Bell Adjustment Inventory for College women. *J. soc. Psychol.*, 1943, **17**, 103–109.

203. Marsh, C. J. The prognostic value of the Washburne Social Adjustment Inventory. *J. soc. Psychol.*, 1943, **17**, 287–294.

204. Marsh, C. J. The influence of supplementary verbal directions upon results obtained with questionnaires. *J. soc. Psychol.*, 1945, **21**, 275–280.

205. Marshall, Helen. Clinical applications on the Bernreuter Personality Inventory. *Psychol. Bull.*, 1933, **30**, 601–602.

206. Martin, H. G. The construction of the Guilford-Martin Inventory of Factors G-A-M-I-N- *J. appl. Psychol.*, 1945, **29**, 298–300.

207. Maslow, A. H., Hirsh, Elisa, Stein, Marcella, & Honigmann, Irma. A clinically derived test for measuring psychological security-insecurity. *J. gen. Psychol.*, 1945, **33**, 21–41.

208. Mathews, C. O. Factors which contribute to undesirable overt behavior. *J. consult. Psychol.*, 1937, **1**, 41–48.

209. Mathews, Ellen. A study of emotional stability in children by means of a questionnaire. *J. Delinqu.*, 1923, **8**, 1–39.

210. May, M. A. Problems of measuring character and personality. *J. soc. Psychol.*, 1932, **3**, 131–145.

211. May, M. A., Hartshorne, H., & Welty, Ruth E. Personality and character tests. *Psychol. Bull.*, 1927, **24**, 418–435.

212. May, M. A., Hartshorne, H., & Welty, Ruth E. Personality and character tests. *Psychol. Bull.*, 1928, **25**, 427–443.

213. May, M. A., Hartshorne, H., & Welty, Ruth E. Personality and character tests. *Psychol. Bull.*, 1930 **27**, 485–494.

214. Meehl, P. E. The dynamics of "structured" personality tests. *J. Clin. Psychol.*, 1945, **1**, 296–303.

215. Michael, J. C. & Buhler, C. Experiences with personality testing in a neuropsychiatric department of a public general hospital. *Dis. nerv. System.* 1945, **6**, 205–211.

216. Miles, D. W., Wilkins, W. L., Lester, D. W., & Hutchens, W. H. The efficiency of a high speed screening procedure in detecting the neuropsychiatrically unfit at a U. S. Marine Corps recruit training depot. *J. Psychol.*, 1946, **21**, 243–268.

217. Mittelmann, B. The Cornell Selectee Index: short form to be used at induction, at reception and during hospitalization. *War Psychiatry: Proceedings of the Second Brief Psychotherapy Council.* Chicago: Institute for Psychoanalysis, 1944.

218. Moore, H. & Steele, Isabel. Personality tests. *J. abnorm. soc. Psychol.*, 1934, **29**, 45–52.

219. Morgan, T. F. A brief study of the

validity of a neurotic inventory. *J. appl. Psychol.*, 1935, **19**, 180–188.

220. MOSIER, C. I. A factor analysis of certain neurotic symptoms. *Psychometrika*, 1937, **2**, 263–286.
221. MOSIER, C. I. On the validity of neurotic questionnaires. *J. soc. Psychol.*, 1938, **1**, 3–16.
222. MURRAY, MARGARET E. Validation of items on the Psychoneurotic Inventory. *J. juv. Res.*, 1932, **16**, 213–230.
223. MURPHY, LOIS B. Personality tests for clinical use. *J. abnorm. soc. Psychol.*, 1932, **27**, 168–171.
224. NACCARTI, S. & GARRETT, H. E. The relation of morphology to temperament. *J. abnorm. soc. Psychol.*, 1924, **19**, 254–263.
225. NEPRASH, J. A. The reliability of questions in the Thurstone Personality Schedule. *J. soc. Psychol.*, 1936, **7**, 239–244.
226. NEYMANN, C. A. & KOHLSTEDT, K. D. A new diagnostic test for introversion-extroversion. *J. abnorm. soc. Psychol.*, 1929, **23**, 482–487.
227. OLIVER, R. A. C. The traits of extraverts and introverts. *J. soc. Psychol.*, 1930, **1**, 345–366.
228. OLSON, W. C. The waiver of signatures in personal reports. *J. appl. Psychol.*, 1933, **20**, 442–450.
229. PAGE, H. E. Detecting psychoneurotic tendencies in army personnel. *Psychol. Bull.*, 1945, **42**, 645–658.
230. PAGE, J., LANDIS, C., & KATZ, S. E., Schizophrenic traits in the functional psychoses and in normal individuals. *Amer. J. Psychiat.*, 1934, **13**, 1213–1225.
231. PAI, T., SUNG, S. M., & HSU, E. H. The application of Thurstone's Personality Schedule to Chinese subjects. *J. soc. Psychol.*, 1937, **8**, 47–72.
232. PALLISTER, HELEN. The negative or withdrawal attitude: a study in personality organization. *Arch. Psychol.*, N. Y., 1938, No. 15.
233. PAPURT, M. J. A study of the Woodworth Psychoneurotic Inventory with suggested revision. *J. abnorm. soc. Psychol.*, 1930, **25**, 335–352.
234. PEDERSEN, RUTH A. Validity of the Bell Adjustment Inventory when applied to college women. *J. Psychol.*, 1940, **9**, 227–236.
235. PERRY, R. C. Analysis of group factors in certain adjustment questionnaires. *Psychol. Bull.*, 1933, **30**, 598.
236. PERRY, R. C. A group factor analysis of the adjustment questionnaire. *Southern California Educational Monogr.*, 1933–1934, **5**, 1–93.
237. PESCOR, M. J. The Woodworth P.D. Sheet as applied to delinquents. *J. Clin. Psychopath.*, 1945, **7**, 383–393.
238. PETERS, C. C. Validity of personality inventories studied by a "Guess Who" technique. *Psychol. Bull.*, 1940, **37**, 453.
239. PINTNER, R. Neurotic tendency and its relation to some other mental traits. *Sch. & Soc.*, 1932, **36**, 765–766.
240. PINTNER, R. & FORLANO, G. Four tests of a personality inventory. *J. educ. Psychol.*, 1938, **29**, 93–100.
241. PINTNER, R. & FORLANO, G. Validation of personality tests by outstanding characteristics of pupils. *J. educ. Psychol.*, 1939, **30**, 25–31.
242. PINTNER, R., LOFTUS, J. J., FORLANO, G., & ALSTER, B. *Aspects of Personality.* Yonkers-on-Hudson, N. Y.: World Book Company, 1937.
243. POE ENG YU. A personality (Bernreuter) study of two groups of psychotics. *Psychol. Bull.*, 1934, **31**, 586–587.
244. REED, H. B. The place of the Bernreuter Personality Schedule, Stanquest Mechanical Aptitude and Thurstone Vocational Interest tests in college entrance examinations.

J. appl. Psychol., 1941, **25**, 528–534.

245. Reed, P. H. & Wittman, Phyllis. "Blind" diagnoses on several personality questionnaires checked with each other and the psychiatric diagnosis. *Psychol. Bull.*, 1942, **39**, 592.

246. Remners, H. W., Whistler, L. & Duwald, V. H. Neurotic indicators at the adolescent level. *J. soc. Psychol.*, 1938, **9**, 17–24.

247. Reusser, J. L. Personal attitudes in delinquent boys. *J. juv. Res.*, 1933, **17**, 19–34.

248. Richmond, Winifred V. *Personality: its development and hygiene.* New York: Farrar & Rinehart, 1937.

249. Riggs, Winifred C. & Joyal, A. E. The validation of the Loofbourow-Keys Personal Index. *J. educ. Psychol.*, 1938, **29**, 194–201.

250. Roback, A. A. *A bibliography of character and personality.* Cambridge, Mass.: Sci-Art Publishers, 1927.

251. Roback, A. E. Personality tests—whither? *Char. & Pers.*, 1933, **1**, 214–224.

252. Roberts, C. S. & Fisher, V. E. Another attempt at measures of extroversion and introversion. *Psychol. Clinics*, 1933, **22**, 88–93.

253. Robertson, Arrie E. & Stromberg, E. L. The agreement between associate's ratings and self-ratings of personality. *Sch. & Soc.*, 1939, **50**, 126–127.

254. Rogers, C. R. *A test of personality adjustment.* New York: Association Press, 1931.

255. Root, A. I. A short test of introversion-extroversion. *Person. J.*, 1931, **10**, 250–253.

256. Root, A. I. *Introversion-Extroversion Inventory.* New York: Psychological Corp., 1932.

257. Root, A. I. & Root, Elizabeth B. A study of the Neymann-Kohlstedt diagnostic test for introversion-extroversion. *J. abnorm. soc. Psychol.*, 1932, **26**, 415–421.

258. Rosenwald, A. The California Personality Scale as a diagnostic instrument. *Psychol. Bull.*, 1942, **39**, 599.

259. Rosenzweig, S. A basis for the improvement of personality tests, with special reference to the Masculinity-Femininity battery. *J. abnorm. soc. Psychol.*, 1932, **26**, 415–421.

260. Roslow, S. Nation-wide and local validation of the P.Q. or Personality Quotient test. *J. appl. Psychol.*, 1940, **24**, 529–539.

261. Roslow, S. The nation-wide validation of a personality test. *Psychol. Bull.*, 1940, **37**, 520.

262. Ruch, F. L. A technique for detecting attempts to fake performance on the self-inventory type of personality test. In *Studies in personality contributed in honor of Lewis M. Terman.* New York: McGraw-Hill, 1942.

263. Ruch, F. L. Ability of adults to fake desirable responses on two personality self-inventories and an attempt to develop a "lie detector" test. *Psychol. Bull.*, 1945, **42**, 539–540.

264. Ruggles, R. & Allport, G. W. Recent applications of the A-S Reaction study. *J. abnorm. soc. Psychol.*, 1939, **34**, 518–528.

265. Rundquist, E. A. Form of statement in personality measurement. *J. educ. Psychol.*, 1940, **31**, 134–147.

266. Rundquist, E. A. & Sletto, R. F. *Personality in the depression: a study in the measurement of attitudes.* Minneapolis: University of Minnesota Press, 1936.

267. Runner, Jessie R. & Seaver, Margaret. A personality analysis test. *Amer. J. Sociol.*, 1943, **49**, 209–222.

268. Ryans, D. G. & Peters, E. F. An analysis and adaptation of the Bell

School Inventory with respect to student adjustment in a women's college. *J. appl. Psychol.*, 1940, **24**, 455–462.

269. Sailer, R. C. & Luh, C. W. Self-estimation of Chinese students. *J. soc. Psychol.*, 1933, **4**, 245–249.

270. St. Clair, W. F. & Seegers, J. C. Certain aspects of the validity of the Bernreuter personality inventory. *J. educ. Psychol.*, 1937, **28**, 530–531.

271. Sargent, Helen. An experimental application of projective principles to a paper and pencil personality test. *Psychol. Monogr.*, 1944, **5**, 1–57.

272. Schmidt, C. H. O. Test profiles as a diagnostic aid: the Minnesota Multiphasic Inventory. *J. appl. Psychol.*, 1945, **29**, 115–131.

273. Schott, E. L. Personality tests in clinical practice. *Psychol. Bull.*, 1935, **32**, 733–734; *J. abnorm. soc. Psychol.*, 1937, **32**: 236–239.

274. Schwegler, R. A. *A study of introvert-extravert responses to certain test situations.* New York: Teachers College Bureau of Publications, Columbia University, No. 361, 1929.

275. Seibert, E. W. An evaluation of the Johnson Temperament Analysis *J. Clin. Psychol.*, 1945, **1**, 193–195.

276. Shen, E. Validity of self-estimate. *J. educ. Psychol.*, 1925, **16**, 104–107.

277. Shen, E. Differences between Chinese and American reaction to the Bernreuter Personality Inventory. *J. soc. Psychol.*, 1936, **7**, 471–474.

278. Shlaudeman, K. W. A new scale for scoring the Bernreuter Personality Inventory. *J. soc. Psychol.*, 1936, **7**, 483–486.

279. Simpson, R. M. Self-rating of prisoners compared with that of college students. *J. soc. Psychol.*, 1933, **4**, 464–467.

280. Slater, P. The psychometric differentiation of neurotic from normal men. *Brit. J. med. Psychol.*, 1945, **20**, 277–279.

281. Slawson, J. Psychoneurotic responses of delinquent boys. *J. abnorm. soc. Psychol.*, 1925, **20**, 261–281.

282. Slawson, J. *The delinquent boy.* Boston: Badger, 1926.

283. Smith, Hattie N. A study of the neurotic tendencies shown in dementia praecox and manic depressive insanity. *J. soc. Psychol.*, 1933, **4**, 116–128.

284. Smith, M. A note on stability in questionnaire responses. *Amer. J. Sociol.*, 1938, **38**, 713–720.

285. Smith, R. B. The development of an inventory for the measurement of the inferiority feelings at the high school level. *Arch. Psychol.*, N. Y. 1932, No. 144.

286. Smith, R. B. The measurement of inferiority feelings at the high school level. *Psychol. Bull.*, 1932, **29**, 633–634.

287. Snyder, Marguerite. A comparison of mental traits and attitudes of delinquent boys and girls. *J. juv. Res.*, 1931, **15**, 181–191.

288. Snyder, W. U. A survey of recent studies in the measurement of personality, attitudes and interests of adolescents. *J. gen. Psychol.*, 1941, **25**, 403–420.

289. Speer, G. S. The use of the Bernreuter Personality Inventory as an aid in the prediction of behavior problems. *J. juv. Res.*, 1936, **20**, 65–69.

290. Spencer, D. The frankness of subjects on personality measures. *J. educ. Psychol.*, 1938, **29**, 26–35.

291. Spencer, D. *Fulcra of conflict.* Yonkers-on-Hudson, N. Y.: World Book Co., 1938.

292. Sperling, A. A comparison of the Human Behavior Inventory with two other personality inventories. *Educ. Psychol. Measmt.*, 1942, **2**, 291–297.

293. STAGNER, R. The intercorrelations of some standard personality tests. *J. appl. Psychol.*, 1932, **16**, 453–464.

294. STAGNER, R. Improved norms for four personality tests. *Amer. J. Psychol.*, 1933, **45**, 303–307.

295. STAGNER, R. Validity and reliability of the Bernreuter Personality Inventory. *J. abnorm. soc. Psychol.*, 1934, **28**, 413–418.

296. STAGNER, R. The Wisconsin Scale of Personality Traits. *J. abnorm. soc. Psychol.*, 1937, **31**, 463–471.

297. STOGDILL, EMILY L. & Thomas, MINNIE E. The Bernreuter Personality Inventory as a measure of student adjustment. *J. soc. Psychol.*, 1938, **9**, 299–315.

298. SUPER, D. E. The Bernreuter Personality Inventory. *Psychol. Bull.*, 1942, **39**, 94–125.

299. SUPER, D. E. Strong's Vocational Interests of Men and Women: a special review. *Psychol. Bull.*, 1945, **42**, 359–370.

300. SWEET, L. *The measurement of personal attitudes in younger boys.* New York: Association Press, 1929.

301. SYMONDS, P. M. *Diagnosing personality and conduct.* New York: Appleton-Century, 1931.

302. SYMONDS, P. M. Measuring the personality adjustment of high school pupils. *Psychol. Bull.*, 1933, **30**, 664–665.

303. SYMONDS, P. M. *Psychological diagnosis in social adjustment.* New York: American Book Co., 1934.

304. SYMONDS, P. M. & BLOCK, VIRRINIA L. *Student Questionnaire.* New York: Bureau of Publications, Teachers College, Columbia University, 1932.

305. SYMONDS, P. M. & JACKSON, C. E. An adjustment survey. *J. educ. Res.*, 1930, **21**, 321–330.

306. SYMONDS, P. M. & JACKSON, C. E. *Measurement of the personality adjustment of high school pupils.* New York: Bureau of Publications, Teachers College, Columbia University, 1935.

307. TERMAN, L. M. The measurement of personality. *Science*, 1934, **80**, 605–608.

308. TERMAN, L. M. & MILES, CATHARINE C. *Attitude-Interest Analysis Test (Masculinity-Feminity Test)* New York: McGraw-Hill, 1936.

309. THOMAS, F. C. Intuition or psychometry in the study of personality? *Character & Pers.*, 1939, **7**, 309–317.

310. THOMSON, W. A. *An evaluation of the P.Q. (Personality Quotient) Test. Character & Pers.*, 1938, **6**, 274–292.

311. THORPE, L. P. *Psychological foundations of personality.* New York: McGraw-Hill, 1938.

312. THURSTONE, L. L. The vectors of the mind. *Psychol. Rev.*, 1934, **41**, 1–32.

313. THURSTONE, L. L. & THURSTONE, THELMA G. *Personality Schedule.* Chicago: University of Chicago Press, 1929.

314. THURSTONE, L. L. & THURSTONE, THELMA. A neurotic inventory. *J. soc. Psychol.*, 1930, **1**, 3–30.

315. TIEGS, E. W., CLARK, W. W., & THORPE, L. P. *California Test of Personality.* Los Angeles: California Test Bureau, 1942.

316. TIEGS, E. W. CLARK, W. W., & THORPE, L. P. The California test of personality. *J. educ. Res.*, 1941, **35**, 102–108.

317. TRAVIS, R. C. The measurement of fundamental character traits by a new diagnostic test. *J. abnorm. soc. Psychol.*, 1924, **19**, 400–420.

318. TRAXLER, A. E. Current construction and evaluation of personality and character tests. *Rev. educ. Res.*, 1941, **11**, 57–79.

319. TRAXLER, A. E. The reliability of the Bell Inventories and their correlation with teacher judgment. *J. appl. Psychol.*, 1941, **25**, 672–678.

320. Traxler, A. E. *The use of tests and rating devices in the appraisal of personality.* New York: Educational Records Bureau, 1942. Educational Research Bulletin No. 23, revised ed.
321. Traxler, A. E. Current construction and evaluation of personality tests. *Rev. educ. Res.*, 1944, **14**, 55–66.
322. Traxler, A. E. *Techniques of Guidance.* New York: Harper, 1945.
323. Tubbs, W. R. A study of the interrelationships between the Adams-Lepley Personal Audit and the Bernreuter Personality Inventory. *J. appl. Psychol.*, 1942, **26**, 338–351.
324. Turney, A. H. & Collins, F. I. An experiment in improving the personality of high school seniors. *J. educ. Psychol.*, 1940, **31**, 550–553.
325. Turney, A. H. & Fee, Mary. An attempt to use the Bell Adjustment Inventory for high school guidance. *Sch. Rev.*, 1936, **41**, 193–198.
326. Tyler, H. T. Evaluation of the Bell Adjustment Inventory. *Junior College J.*, 1936, **6**, 353–357.
327. van Vorst, R. B. An evaluation of test performances of a group of psychopathic delinquents. *Psychol. Bull.*, 1943, **50**, 583.
328. van Wagenen, N. B. *The stability of self-description tests of personality adjustment.* New York: Author, 1935.
329. Vernon, P. E. The attitudes of the subject in personality testing. *J. appl. Psychol.*, 1934, **18**, 165–167.
330. Wang, C. K. A. The internal consistency of Allport's Ascendance-Submission Test. *J. abnorm. soc. Psychol.*, 1931, **26**, 154–156.
331. Warner, N. & Gallico, Margaret W. Cornell Service Index. Report on its use in the evaluation of psychiatric problems in a naval hospital. *War Med.*, 1945, **7**, 214–217.
332. Washburne, J. N. A test of social adjustment. *J. appl. Psychol.*, 1935, **19**, 125–144.
333. Washburne, J. N. *Washburne Social-Adjustment Inventory.* Yonkers-on-Hudson, N. Y.: World Book Co., 1940.
334. Wasson, Margaret M. The agreements among certain types of personality schedules. *J. Psychol.*, 1940, **9**, 351–363.
335. Watson, G. Measures of character and personality. *Psychol. Bull.*, 1932, **29**, 147–176.
336. Watson, G. Character and personality tests. *Psychol. Bull.*, 1933, **30**, 467–487.
337. Watson, G. Next steps in personality measurement. *Character & Pers.*, 1933, **2**, 66–73.
338. Watson, G. Mental hygiene and emotional adjustment. *Rev. educ. Res.*, 1935, **5**, 245–258.
339. Watson, G. The specific techniques of investigation: testing intelligence, aptitudes, and personality. *Yearbook. Nat. Soc. Stud. Educ.*, 1938, **3**(**II**), 357–373.
340. Watson, R. I. The relationship of the Affective Tolerance Inventory to other personality inventories. *Educ. Psychol. Msmt.*, 1942, **2**, 83–90.
341. Watson, R. I. Clinical validity of the Inventory of Affective Tolerance. *J. soc. Psychol.*, 1945, **22**, 3–15.
342. Watson, R. I. & Fisher, V. E. An Inventory of Affective Potency. *J. Psychol.*, 1941, **12**, 139–149.
343. Weider, A., Brodman, K., Mittelmann, B., Wechsler, D. & Wolff, H. G. Cornell Service Index. *War Med.*, 1945, **7**, 209–213.
344. Whitman, R. H. A short scale for measuring introversion-extroversion. *J. appl. Psychol.*, 1929, **13**, 499–504.
345. Wile, I. S. Chance in the validation of personality tests. *Amer. J. Orthopsychiat.*, 1942, **12**, 154–160.

346. WILEY, L. N. & TRIMBLE, O. C. The ordinary objective test as a possible criterion of certain personality traits. *Sch. & Soc.*, 1936, **43**, 446–448.

347. WILLIAMSON, E. G. & DARLEY, J. G. *The Minnesota Inventory of Social Preferences.* New York: The Psychological Corp., 1937.

348. WILLOUGHBY, R. R. Some properties of the Thurstone personality schedule and a suggested revision. *J. soc. Psychol.*, 1932, **3**, 420–424.

349. WILLOUGHBY, R. R. *Willoughby (Clark-Thurstone) Personality Schedule.* Riverside, R. I.: Author, 1932.

350. WILLOUGHBY, R. R. Norms for the Clark-Thurstone Inventory. *J. soc. Psychol.*, 1934, **5**, 91–97.

351. WITTMAN, MARY P. & HUFFMAN, A. V. A comparative study of developmental, adjustment, and personality characteristics of psychotic, psychoneurotic, delinquent, and normally adjusted teen aged youths. *J. genet. Psychol.*, 1945, **66**, 167–182.

352. WOLFLE, D. Factor analysis in the study of personality. *J. abnorm. soc. Psychol.*, 1942, **37**, 393–397.

353. WOODWORTH, R. *Personal Data Sheet.* Chicago: Stoelting, 1919.

354. WRIGHTSON, J. W. Validity of the Woodworth-Mathews P.D. Sheet for diagnosing certain personality disorders. *J. educ Psychol.*, *1934*, **25**, 39–44.

355. YOURMAN, J. Children identified by their teachers as problems. *J. educ. Sociol.*, 1932, **5**, 334–343.

356. ZAPE, ROSALIND M. Comparisons of responses to superstitions on a written test and in actual situations. *J. educ. Res.*, 1945, **39**, 13–24.

357. ZUBIN, J. The method of internal consistency for selecting test items. *J. educ Psychol.*, 1934, **25**, 345–356.

358. *Character Education.* Report of the Committee on Character Education of the National Education Association. Washington: Government Printing Office, 1926.

359. *Character Education.* Tenth Yearbook of the Department of Superintendence, National Education Association. Washington: National Education Association, 1932.

360. *Cornell Service Index.* New York: Cornell University Medical College, 1945.

THE PRESENT STATUS OF PSYCHOLOGY IN SOUTH AMERICA*

MARGARET E. HALL
Bureau of Child Study
Chicago Board of Education

It is important to realize at the outset that psychology as an independent *profession* can still hardly be said to exist in South America. Psychology is taught in universities and normal schools, even at times in secondary schools and it is applied in a multitude of practical ways but the people who serve psychology come from other professional fields and have not usually had the type of academic preparation which in the United States and Europe we tend to consider requisite. The people now teaching and practicing psychology come mainly from three professional fields, medicine, pedagogy and philosophy. These fields determine their basic training and continue later to determine their professional orientation. One still does not see students entering universities with the idea of becoming "psychologists" and, indeed even should they have such a professional aim, there would be little opportunity to obtain the necessary academic training as universities in Latin America are now constituted. For that reason a student must choose another and primary career and work into psychology later through personal reading, apprenticeship or study in a foreign country. People frequently enter psychological work or teaching after years of medical practice, university teaching in other fields, school administration or elementary or secondary school teaching. Defined in terms of the specialized professional study considered necessary in the United States for entering the career of psychology, there can be said to be almost no "psychologists" in South America, aside from a few European psychologists now working there or the rare few among South Americans who have had psychological training abroad. Most South Americans now working in psychology, although often highly trained in other fields, are seldom professionally trained in psychology, as such.

---

* The author spent the entire year 1944–1945 in South America visiting all ten of the countries on that continent. In each country contact was made with nearly all persons reported to be engaged in either the teaching or practice of psychology as well as with agencies concerned with "psychological" work although the personnel might not have included "psychologists." Contacts were greatly facilitated by the Department of State in Washington which instructed the cultural attachés of the American Embassies to assist the author. The cultural attachés were very helpful both in suggesting profitable contacts for obtaining information about psychological work and frequently in making specific arrangements for visits.

The author wishes particularly to express her gratitude and appreciation to the many South American colleagues who gave so generously of their time and received her with a cordiality and friendship far transcending any formal expression of "good neighborliness."

The following discussion of psychology in South America is organized by countries. Within the account of the work in each country there is first a discussion of the academic psychology offered, followed by a discussion of the applied fields, with the latter given the greater emphasis in this article. The reason for this emphasis is two-fold. First, there was a very complete and able presentation of academic psychology and the history of psychology in South America* by Beebe-Center and McFarland published in this journal in 1941, pp. 627–667. In the second place the present writer finds little continuity to the academic work offered in most South American universities. Many of the professors listed a few years ago as teaching psychology are no longer in the field at all or are in other psychological activities, with new professors teaching the courses. The reason for this may be that many of those who have taught or are now teaching psychology have an only incidental or transitory interest in the subject. Courses are not infrequently taught by a different professor each year or two. Then, too, many South American universities are in a state of change and reorganization and, in any case, are more apt than our universities to reflect political change and turmoil. Professors are not uncommonly dismissed because of their political opinions and in general have less security of tenure than is usual in the United States.

PSYCHOLOGY IN THE INDIVIDUAL COUNTRIES

ARGENTINA

Academic psychology. At the University of Cuyo in Mendoza, Argentina, there is the Instituto de Psicología Experimental under the direction of Professor Horacio J. A. Rimoldi. Dr. Rimoldi spent several years in the English universities, studying with Spearman and others, following his medical training in Argentina. His chief interest is in factor analysis and he is carrying out a number of factorial research studies in addition to teaching courses in psychology. A number of studies have already been published.

At the University of Córdoba in Córdoba, Argentina, only one course is offered in psychology, taught by a professor of philosophy, Francisco W. Torres. Prof. Torres has published a number of metaphysical studies and is now preparing a book on systematic psychology based on the metaphysics of Aristotle. At the Colegio Monserrat, a secondary school for boys, in connection with the University of Córdoba, a course in psychology is required in the last year and is taught by Prof. Nestor A. Pizarro who has prepared his own text for the course.

---

* BEEBE-CENTER, J. G. & MCFARLAND, R. A., Psychology in South America, *Psychol. Bull.* 1941, **38** 627–667.

The University of Buenos Aires has had traditionally more facilities in psychology than other universities in Argentina although at present changes are in progress and little psychology is being offered. In the Faculty of Philosophy and Letters there is an Institute of Psychology, formerly under Prof. Enrique Mouchet, an outstanding contributor to psychology. Prof. Mouchet has now retired and his place has not as yet been filled. Prof. José L. Alberti, a medical doctor, is the chief of the Laboratory of Experimental Psychology in this institute and teaches the course in experimental psychology. A course in general psychology follows the experimental course and is taught by Prof. Coriolano Alberini, a philosopher. The University of Buenos Aires has perhaps the outstanding department of psychiatry in South America. It is headed by Dr. Gonzalo Bosch who is also director of the Argentine League of Mental Hygiene (Liga Argentina de Higiene Mental). All medical students are required to take one course in psychiatry and there are additional courses for those specializing in the field. In addition to the regular courses in psychiatry a course in medical psychology is offered by Dr. Eduardo Krapf, a German psychiatrist now residing in Buenos Aires. The department of psychiatry has its own building with extensive laboratories for research and can draw its cases from an immense mental hospital within whose grounds it is located.

Buenos Aires is the only center in South America where a large and active psychoanalytic group is to be found. The group is organized under the name of the Asociación Psicoanalítica Argentina and its members are all either medical or lay analysts in practice. The director at the time of the writer's visit was Dr. Enrique J. Pichon-Rivière who is also a professor of psychiatry at the University of Buenos Aires. The group meet for study and discussion three times a week and publish a journal, the *Revista de Psicoanálisis*. The Asociación has already become a center of training in psychoanalysis, with psychiatrists from other countries, principally Brazil, coming to study. It has plans for expansion, with a building of its own and better facilities for research and training. One leading psychiatrist told the writer that among the approximately 120 psychiatrists in Argentina, about twenty are psychoanalytic.

Dr. Alfredo D. Calcagno of the University of La Plata, in La Plata, Argentina, is generally regarded as the leading authority in the country on educational psychology. He has a laboratory of experimental psychology and teaches psychopedagogy in the Faculty of Humanities and Educational Sciences. Dr. Calcagno was recently appointed rector of the University. Other professors of psychology in the university are Ernesto L. Figueroa, F. Romero and Eugenio Pucciarelli.

The University of the Coast (Universidad del Litoral) is divided in location, with most of the faculties in Rosario, but some in Santa Fé and Paraná. Little psychology is taught but psychiatric training is offered in the medical faculty under Antonio Foz and Teodoro Fracassi.

Normal schools in Argentina, as in other South American countries, usually offer from one to three courses in psychology, frequently taught by professors of philosophy or education or by medical doctors. Secondary schools (colegios or liceos) also offer one course in psychology. Schools of social service offer one or two courses in psychology which are usually taught by psychiatrists.

Applied psychology. Argentina has very little to offer at the present time in the various fields of applied psychology. The national Ministry of Justice and Public Instruction has no psychological program whatsoever. There are no facilities for child study or even mental measurement within the structure of the educational system itself and the resources for care of deviate children are very limited. There are two children's institutions which have some psychiatric and psychological service, the Hogar "Santa Rosa" for delinquent girls committed by the Juvenile court and the Escuela de Adaptación for delinquent, mentally defective or academically retarded boys. Dr. Carolina Tobar-Garcia, a psychiatrist, serves both institutions and has an assistant who gives mental tests. There are no special classes within the public schools themselves for special training of exceptional children.

Buenos Aires has one of the most important child guidance clinics in South America, the Centro de Psicología y Psiquiatría Infantil, which is under the Institute of Pediatrics of the Clinical Hospital of the University of Buenos Aires. The director of the clinic is Dr. Telma Reca de Acosta, a pediatrician who has studied child psychiatry and has had some training in the United States. There are about ten members on her staff, all on a part-time, volunteer basis and including other psychiatrists, an educator, Nicolas M. Tavella, who gives the psychological tests and several social workers. The clinic operates along the usual lines of child guidance clinics, accepts children up to 16 years of age and with a wide variety of problems and from many sources of referral. Treatment is available in a limited number of cases. The clinic also functions as a training center for student psychiatrists and other child guidance workers.

The Argentine Social Museum (Museo Social Argentino) which is an important adult educational and cultural organization in Buenos Aires has as one of its subsidiary divisions the Institute of Professional Guidance (Instituto de Orientación Profesional). The director is Dr. Gregorio Fingerman and the function of the Institute is vocational guidance. There are ten part-time assistants who have been trained by Dr. Fingerman. The project was started over twenty five years ago when a group of German psycho-technicians were brought to Argentina under a government contract to initiate psycho-technical research and teach at the Instituto Nacional del Profesorado Secundario. Dr. Fingerman received his training from them and continued the program after their return to Germany. The Instituto Nacional del Profesorado Secundario no longer offers psychological work but the applied work goes on in the Museo Social Argentino. The vocational guidance services are used by private individuals, by schools and by various government agencies such as the police and various branches of the Army and Navy, which formerly had their own psychological laboratories.

There are a few small organizations in Buenos Aires offering various types of psychological service or training, the Instituto "Sigmund Freud" directed by Bela Székely and the Centro de Psicología under Sra. de Aparicio. Dr. Nestor Vellejos Meana, a medical doctor connected with the Sociedad de Beneficencia de la Capital, is interested in psychopedagogy and does some work with mental tests in connection with school medical service.

The Argentine League of Mental Hygiene, under Dr. Gonzalo Bosch, has a social service department and the Instituto Neuropsiquiátrico within it and publishes the *Revista Argentina de Higiene Mental.*

Within the medical faculty of the Universidad del Litoral in Rosairo the psychiatry department has one section for children. They are housed in a small building separate from adult patients and may stay indefinitely. All types of deviate children seem to be accepted, from mental defectives or deaf to emotionally disturbed children. Some speech therapy is provided for children needing such care.

The Psychiatric Hospital of the Province of Santa Fé and located in the city of Santa Fé is outstanding and is the most modern mental hospital building in Argentina. The director is Dr. Luis E. Bonsembiante, who also teaches a course in psychopathology in the provincial school of social service.

BOLIVIA

Academic psychology. Very little psychology, if any, is taught in the six universities of Bolivia. At the Escuela Nacional de Maestros in Sucre, Bolivia, which is the training school for primary and secondary school teachers, several courses in psychology are offered, with Dr. Alfredo Vargas the principle professor of the subject.

Applied psychology. The Pedagogical Department of the Ministry of Education of Bolivia is the only agency engaged in applied work in the country. The Department has two sections, the Section of Measurements, under the direction of Dr. Alfredo Vargas, is located in Sucre and closely affiliated with the Escuela Nacional de Maestros. The other section, the Section of School Plans and Programs, is under the direction of Prof. Guido Villa-Gómez and is located in La Paz, the capital. Its chief function up to the present has been "controlo escolar," the working out of standards of achievement for each grade and the construction of achievement tests on which to base promotion. Prof. Villa-Gómez and his staff hope to extend their service shortly to the psychological study of individual children. Prof. Villa-Gómez had his professional training in Brazil and expects to go to the United States to study within the coming year. Another member of the staff, Assistant Prof. Elvira Pommier, also plans to study in the United States soon.

BRAZIL

Academic psychology. The University of Porto Alegre offers the following three courses in psychology, all taught at the present time by psychiatrists, general psychology, psychopathology and child psychology. The latter is taught by Dr. Decio Soares de Souza who had psychological training at Columbia University in New York in addition to psychiatric training. Dr. de Souza has a laboratory at the psychiatric hospital and is carrying out research on experimental neuroses in animals. He is the professor of psychiatry in the medical faculty of the university.

The city of São Paulo, Brazil, has probably the most extensive facilities for psychological study of any South American city. At the University of São Paulo psychiatry is taught in the Psychiatric Clinic of the medical faculty by Dr. A. C. Pacheco e Silva. A new building to house the Psychiatric Clinic is now under construction, with the course in psychiatry held in the meantime at the Institução de Assistencia Social ao Psicopata. Dr. Pacheco e Silva publishes extensively and was active in the program of selection of men for the Brazilian Expeditionary Forces. The Faculty of Philosophy, Science and Letters has one of the several best centers in South America for the study of educational psychology. Prof. Noemy da Silveira Rudolfer, temporarily on leave to teach in the University of Paraguay, offers several courses in psychology, assisted by Anita Castillo M. Cabral, and heads the Laboratory of Psychology.

There is a unique and interesting private institution in São Paulo, the Escola Livre de Sociología e Política de São Paulo, which, though not a part of the university, is affiliated with it. It is supported in part by funds from the State of São Paulo, in part by funds from the Rockefeller Foundation and in part by private contributions. The list of courses offered there is longer and more varied than that encountered by the writer in any other institution. The following courses are offered:

Introduction to psychology	Cecilia de Castro Silva
Introduction to social psychology	Cecilia de Castro Silva
Psycho-technics	Dr. Robert Mange and Maria de Lourdes Campos Viegas
Mental hygiene	Dr. Durval Marcondes (psychiatrist)
Psychoanalysis	Dr. Durval Marcondes (psychiatrist)
Social service	Dr. Pacheco e Silva (psychiatrist)

Various other courses closely related to psychology are offered. The staff is partly Brazilian and partly North American. Dr. Donald Pierson (M.A. and Ph.D. in sociology, University of Chicago) teaches sociology and anthropology, is very active in research and has translated into Portuguese a number of important United States texts in the social sciences. The staff hope that closer relationships can be developed between this school and research foundations and universities in the United States so that students in the social sciences will be able to go to Brazil to do their research in these fields. The possibilities for research in anthropology and other social sciences are particularly rich in Brazil.

At the University of Brazil in Rio de Janeiro psychology is taught in two sections of the National Faculty of Philosophy. Prof. Manoel Lourenço Filho is professor of psychology in the Pedagogical Section but has been on leave to the Ministry of Education for several years. The sequence of three courses is taught by Prof. Carlos Queiroz. Prof. Nilton Campos offers three courses, general, philosophical and experimental psychology, in the Philosophy Section. Prof. Campos is occupied now in writing a critique of Watsonian behaviorism and gestalt psychology. Also within this section is Prof. Arthur Ramos who teaches three courses, general anthropology, physical anthropology and anthropology of Brazil. Also connected with the University of Brazil is the Institute of Psychology headed by Dr. Jayme Grabois, a psychiatrist. Psychiatry is taught in the Institute of Psychiatry of the Faculty of Medicine by Dr. Henrique Roxo and Dr. Januario B. Hencourt.

At Bennett College, a private primary and secondary school and junior college, Prof. Heloisa Marinho teaches three psychology courses, general and educational, child and adolescent psychology. Prof. Marinho formerly taught a course in educational psychology also at the Instituto de Educacão, the official school of the Federal District for the training of primary teachers. She is now conducting experiments there in methods of teaching beginning reading.

Applied psychology in São Paulo. Within the student health division of the Department of Education of the state of São Paulo there is a mental hygiene section including a child guidance clinic (Clínica de Orientacão Infantil da Seccão de Higiene Mental Escolar). Dr. Durval Marcondes, a psychiatrist with psychoanalytic training, is director of the clinic. The staff includes two additional psychiatrists, several psychologists and psychiatric social workers. Children with behavior or personality problems are referred from the schools and the cases are studied by a team of experts, in the manner which is usual in North American child guidance clinics. Intensive treatment is carried on with a relatively small number of cases.

The Cruzado Pro-Infancia of São Paulo, a free kindergarten project for children of working mothers, has a psychology section headed by Dr. Betti Katzenstein, a German psychologist who had her training with Stern in Hamburg. Dr. Katzenstein spends several mornings a week working with problem children in these kindergartens. The Instituto de Pesquisas Juveneis (Institute of Juvenile Research) carries on psychological work with delinquent and dependent children.

In the field of adult mental hygiene and psychiatric care there is the Institução de Assistencia Social ao Psicopata under the direction of Dr. Paulo de Camargo. This institution is a large clinic belonging to the state of São Paulo for the mental examination and diagnosis of adult cases and for treatment of ambulatory cases. Patients requiring hospitalization or prolonged treatment are sent to the state hospital at Juquerí. About 2,700 cases are seen annually. The institution publishes a monthly bulletin, *Boletim de Higiene Mental.*

Brazil is the only South American country where a good beginning is being made in the field of industrial, personnel and vocational psychology. Because of São Paulo's highly industrial character it is perhaps natural that much of this work should have centered there. São Paulo has two important organizations of this type. The larger organization has a branch in both São Paulo and Rio de Janeiro and is called the Servicio Nacional de Aprendizagem Industrial (National Service of Industrial Apprenticeship) but is commonly called "SENAI" from its initials. The organization is supported jointly by private industries, including the railroads which formerly had their own psychotechnical laboratories, and by a grant from the government. The organization as a whole is primarily concerned with problems of training within industry but one section, the Division of Selection, is psychotechnical in function and staff. The head of the division is Prof. Walter Barioni. Dr. Robert Mange is in charge of psychotechnics and other psychologically trained people on the staff are Dr. Aniela Meyer Ginsberg, a Polish psychologist, Dr. Betti Katzenstein, mentioned above, and Italo Bologna. The primary function of the division is job-analysis and the construction and standardization of trade tests for the classification of workers in training courses within the industries which support SENAI. Adaptations are made of some of the United States and European tests and new tests are also developed by the psychotechnicians of the staff. Batteries have been developed so far for mechanical, carpentry and textile trades. The organization plans to extend its services eventually into vocational guidance.

The other organization of this type in São Paulo is the Instituto de Organizacão Racional do Trabalho (Institute for the Rational Organization of Work), again known by its initials, IDORT. This organization is concerned with promoting the over-all efficiency of management and industrial organization as well as of labor. It is supported by the memberships of individual industrialists and laborers. The board is made up of engineers and industrial owners. The present director, Dr. Moacyr E. Alvaro, is an ophthalmologist who became interested in industrial efficiency through his interest in problems of industrial lighting. The Institute has begun a small program of vocational guidance which it hopes to expand. The two European psychologists on the staff of SENAI, discussed above, Dr. Ginsberg and Dr. Katzenstein, spend part of their time at IDORT examining boys referred for guidance.

Applied psychology in Rio de Janeiro. Four agencies of the national government of Brazil have some type of psychological service connected with them. All are located in Rio de Janeiro and all are found under the Ministry of Education and Health, although unrelated with each other. The National Institute of Pedagogical Studies (Instituto Nacional de Estudos Pedagógicos) is directly under the minister of education and is headed by Dr. Manoel Lourenço Filho who is the outstanding figure in the field of educational psychology in South America. The function of the Institute is fundamental research which can be used as a basis for curriculum construction and the development of educational materials and methods. It is a fact-finding and a fact-interpreting agency rather than administrative or supervisory. Dr. Lourenço and his staff have been engaged for years upon an extensive vocabulary study of all age levels, preschool to adult, and in various parts of the nation. Another function of the Institute, somewhat unrelated to its other functions, is the construction of tests for the selection of federal employees. The Institute publishes the *Revista Brasileira de Estudos Pedagógicos* and also a series of research bulletins.

The Division of Industrial Education (Divisão do Ensino Industrial) of the Ministry of Education and Health is becoming interested in the question of vocational guidance and the scientific selection of students for the nation's industrial schools. A psychologist from the United States, Otto Schlesinger, has been working in the division for over two years on the development of this program including the construction of tests for selection. This division, too, is one of several agencies which have arranged jointly to bring Dr. Emilio Mira y López, a Spanish psychiatrist and psychologist, formerly in Montevideo and Buenos Aires, to Rio to offer training in vocational guidance.

One of the major divisions of the Ministry of Education and Health is the National Children's Department (Departamento Nacional da Criança) which corresponds roughly in function to our United States Children's Bureau in the Department of Labor. The director is Prof. Olinto de Oliveira. Its duties are only incidentally psychological but a member of the staff, Madame Helene Antipoff, is well known as a psychologist and a specialist in the education of exceptional children. She had her professional training in Switzerland and France but has been active professionally in Brazil for a number of years. Until recently she was in Belo Horizonte in charge of an institution for mentally retarded children which was founded by a private organization, the Sociedade Pestalozzi, and later taken over by the state of Minas Gerias. She is now organizing a similar project in Rio de Janeiro. Dr. Gustavo Lessa, a physician on the department staff, is also interested in mental hygiene aspects of child welfare, and Dr. Leme Lópes, a psychiatrist, examines some children there.

One of the divisions of the Department of Health of the Ministry of Education and Health is the National Service for Mental Patients (Servicio Nacional de Doenças Mentais), whose director is Dr. Adauto J. Botelho. In addition to the usual mental hospitals and clinics for adults this division includes the Hospital de Neuropsiquiatria Infantil. The director is Dr. Waldemiro Pires, a neurologist, and the psychologist is Marianne Wohlgemuth-Schreyer, who received her professional training in Switzerland. All types of children are received for examination and many stay for considerable periods of time. Each child receives a psychological examination as part of his admission examination.

The public school system of the Federal District in which the capital, Rio de Janeiro, is situated, has various well organized services of a psychological nature. These come under one of the major divisions, the Center for Educational Research (Centro de Pesquisas Educacionais) whose director, Dr. Pedro Pernambuco Filho, is a psychiatrist. This center has, in turn, four services, three of which are definitely psychological. The Psychology and Orthophrenic Service (Servicio de Psicología e Ortofrenía) is headed by Ofélia Boisson Cardosa who was trained in France. This service operates as a psychological clinic, with children referred there for study from the public schools of Rio. The cases so far tend to be predominantly problems of mental or academic retardation. The Educational Guidance Service (Servicio de Orientação Educacional), headed by Aracy Muniz Freire, is concerned with student counseling in the schools. The Anthropometric Service, with Dr. Bartos d'Avile in charge, is engaged in anthropometric research, particularly problems of racial differences in physical growth. The Service of Measurements and Programs (Servicio de Medidas e Programas) is under Isa Goulart Macedo and is concerned with the construction, standardization and use of achievement tests in the various school subjects at all the grade levels. These tests are used widely in the Rio schools as a basis for homogeneous grouping of pupils.

In Rio de Janeiro, as in São Paulo (see above) there is a branch of the Servicio Nacional de Aprendizagem Industrial or SENAI. Octavio Lins Martins is the director of the Division of Selection of the Rio office. He is said to be much interested in research in factor analysis.

The "Getulio Vargas" Foundation (Funcação "Getulio Vargas") is a very new organization also concerned with industrial training problems. Its work, however, unlike SENAI, is directed toward training of the higher levels of employees, technicians, efficiency engineers, psychologists, for industry. It also hopes to develop a vocational guidance program Dr. J. R. Jubé will teach psychology in this training center.

A psychological laboratory was established some years ago at the aviation training school at Campo Dos Afonsos near Rio for the examination of cadets. Several medical doctors conduct the psychological testing and the laboratory is under the Department of Aviation Health of the Army (Departamento de Saúde Aeronautica do Exercito) whose director is Brigadeiro Médico Angelo Godinho dos Santos. Cadets who fail to make satisfactory adjustment during the course of training are sent back to the laboratory for further diagnostic testing and psychiatric study.

One of the most outstanding private psychiatric clinics and sanitoria seen by the writer in South America is to be found in Rio. It is the Clínica Tijuca and Sanatório da Tijuca directed jointly by Dr. Arruda Camara and his wife, Dr. Iracy Doyle. Both Dr. Camara and Dr. Doyle have had extensive psychiatric training in the United States and both are psychoanalytically oriented.

CHILE

Academic psychology. Of the universities in Chile the most extensive work in psychology is to be found at the University of Chile in Santiago. Courses are offered in several of the faculties of this university. Under the Pedagogical Institute of the Faculty of Philosophy and Education approximately five courses have usually been offered, although this Institute has been undergoing some reorganization recently. Dr. Irma Salas who received the Ph.D. degree from Columbia University in New York has taught general psychology, educational psychology and mental measurements and Prof. Arturo Piga has taught child and adolescent psychology. The primary function of the Institute is the preparation of secondary and normal school teachers. Within the Pedagogical Institute is an Institute of Psychology which has a research function primarily. It is headed by Prof. Abelardo Iturriaga who is working on the translation, adaptation and standardization on local populations of various foreign tests, including United States tests. The laboratory is available to anyone who wishes to use it for research. A publication, *Archivos del Instituto de Psicología*, is issued from time to time with research studies from the Institute. A course in the psychology of adolescence is also taught in the School of Fine and Applied Arts of the University.

The psychiatric department of the Faculty of Medicine of the University of Chile occupies a small building within the grounds of the large national mental hospital, Casa de Orates, located in Santiago. The chair of psychiatry, established over thirty years ago and said to be the oldest in South America, is occupied by Dr. Arturo Vivado. There are about six additional professors in the department. All medical students are required to take a full year's course in psychiatry in their last academic year. Psychiatric students go on to more specialized work. Last year a special course in psychoanalysis was given to the staff of the psychiatry department and the mental hospital by Dr. D. Ignacio Matte, a young psychiatrist who has only recently returned to Chile after ten years of psychiatric study and practice in England and the

United States. In this country Dr. Matte studied at the Phipps Clinic of Johns Hopkins University and taught psychiatry at Duke University. Dr. Matte also teaches a course in clinical psychology to the nurses of the Psychiatric Clinic of the Faculty of Medicine and has given a series of lectures to the students of one of the schools of social service in Santiago. An active program of research is going on in the Psychiatric Clinic. One of the staff, who also studied at Johns Hopkins, is carrying on work in experimental neuroses in dogs in relation to drugs.

The University of Concepción in Concepción, Chile, has five courses in psychology offered in the School of Education. All are taught by Prof. Corina Vargas de Medina, who had much of her professional training at Teachers College, Columbia University a number of years ago. She teaches general, educational, child and adolescent psychology and mental and educational measurements. The Catholic University of Chile in Santiago also offers several courses in psychology.

The normal schools for the preparation of urban primary teachers all offer a sequence of three courses in psychology, general, educational and child psychology. The teachers of these courses have usually had only the training provided in the Pedagogical Institute of the University of Chile, although those who teach in the three normal schools of Santiago all happen to have had some additional training in the United States. No psychology is offered in normal schools for the preparation of rural teachers. Psychology is taught in all three of the schools of social service in Santiago. In the Escuela de Servicio Social de la Junta de Beneficencia three courses are required, one taught by a psychiatrist, the other two by psychologists.

Applied psychology. Almost all of the small amount of applied psychological work to be found in Chile is under the various ministries of the national government. The Developmental School (Escuela de Desarrolla) under the Ministry of Education and directed by Juan Sandoval, is an organization for mentally defective children. It has several divisions of work, a small day school in Santiago for mentally retarded boys and girls, a few special classes for retarded children, located in the regular primary schools and two small residence institutions at the outskirts of Santiago for the more retarded cases. The staff of the Escuela de Desarrolla is made up of teachers and social workers who use psychological tests in the examination of children for placement. The intelligence test most used is a Chilean translation and standardization of the 1916 Terman Revision of the Binet scale by Luis A. Tirapegui. The Porteus Mazes are used some and several European tests.

A few children are studied psychologically by Prof. Abelardo Iturriaga in the Instituto de Psicología of the University of Chile (see above) and there is a small department of psychosomatic medicine in the children's hospital, but there is no psychological service of any extent in Santiago to which all children are eligible.

Under the Primary Department of the Ministry of Education the

Institute of Pedagogical Research is carrying on a very active program of practical research in educational methods and materials at the elementary school level. Studies are carried out in eleven experimental primary schools under the direction of Prof. Alfonso Aguirre, head of the Institute, by the staff which numbers about eight. Among the studies made have been adaptations of American educational tests.

There are two institutions for delinquent or dependent boys under the maternal and child welfare division of the Ministry of Health, Social Welfare and Assistance. The Casa de Menores is a small institution in Santiago to which delinquent boys are committed by the Juvenile court. Each boy is examined by the psychiatrist, Dr. Eduardo Brücher Encina and by a psychological worker, Manuel Zamorano, and his case is investigated by one of several social workers. The other institution is the Polytechnical School located in the country some miles from Santiago to which dependent, non-delinquent or mildly delinquent boys are sent. Various types of vocational and agricultural training are available. Some educational and vocational guidance and occasional psychological study of individuals are done.

In a privately financed institution for dependent children, the Children's City (Ciudad del Niño) a Spanish psychologist, Matilde Huici, does individual psychological studies of children.

COLOMBIA

Academic psychology. Colombia is surprisingly lacking in both academic and applied psychology. Very few courses are offered in the several universities of the country and only a very small amount in the urban normal schools, none in the rural normal schools. In the Escuela Normal Superior, a training school for secondary school teachers, in Bogotá, there is a small laboratory of psychology under Prof. Luis Alejandro Vargas where mental testing is done, and Dr. Francisco Socarras, a psychiatrist, does some teaching.

Applied psychology. There is one very outstanding project in applied psychology going on in Colombia. Dr. Mercedes Rodrigo, one of the leading psychologists of Spain, established the Secciõn de Psicotécnia in 1939 in the Faculty of Medicine of the National University of Colombia in Bogotá. She began that year to construct a battery of tests for the selection of students applying for admission to the medical school. As a result of the success of her initial project entrance examinations became compulsory for all medical students and, within the few years since, have been extended to all the other faculties of the university. Dr. Rodrigo's assistant, also trained in Spain, is José García Madrid. The battery of tests is now being extended downward to lower age levels. Dr. Rodrigo does some individual counseling of students who seek her help and she is hoping to initiate some work with children as well.

Colombia has no outstanding training center for psychiatry nor any well known psychiatric clinic. The psychiatrists in private practice tend to be of traditional schools of psychiatric thought and there is little interest so far in psychoanalysis or others of the more dynamic modern schools.

The educational system has developed practically nothing so far in the way of psychological services and there are no child guidance clinics.

ECUADOR

Academic psychology. So far as could be learned little psychology is taught in the four universities of Ecuador. In the Instituto Superior de Pedagogía y Letras in Quito, Prof. Edmundo Carbo is professor of psychology and also under secretary in the Ministry of Education. He constructs psychological tests and trains his students to give them. Prof. Ermel N. Velasco is also associated with this work. There is a laboratory of psychology in a large secondary school for boys in Quito, the Instituto Nacional Mejía, and psychology is taught in the last year by Prof. Gerardo Larrea as part of the work in biological sciences. Psychiatry is taught in the medical faculty of the Central University of Ecuador in Quito by Dr. Julio Endara who is also Vice-Rector of the university. He and Dr. Jorge Escudero are in charge of the psychiatric clinic for adults and are also associated in teaching a course in legal psychiatry in the law faculty.

Applied psychology. The Polyclinic under the Ministry of Social Welfare and Labor has a small service for the psychiatric study of delinquent children. Dr. Jorge Escudero is the psychiatrist and Prof. Gerardo Larrea gives psychological tests.

The American School of Quito has a part-time psychological worker, Carlota de Garcés, who makes some individual case studies of children having educational difficulty.

PARAGUAY

Academic psychology. The Faculty of Humanities of the National University of Paraguay in Asunción has recently instituted some courses in psychology. Prof. Noemy da Silveira Rudolfer of the University of São Paulo, Brazil (see above under Brazil) has been brought to Paraguay for two years to initiate this work. She is teaching courses in general and in educational psychology.

Dr. Carlos Alvarez teaches psychiatry in the Faculty of Medical Sciences and is also director of the Psychiatric Institute of Asunción, a psychiatric clinic and mental hospital. Dr. Andrés Riverola, assistant director, is interested in child guidance and mental hygiene but little work of this type has as yet been started in Paraguay. Prof. María Adela Garcete Speratti teaches general and child psychology at the Normal School in Asunción.

Applied psychology. There are no organizations as yet in Paraguay which are devoted to applied psychology, except for the Psychiatric

Institute of Asunción mentioned above. Several people in Asunción are greatly interested in developing services for children but no facilities are made available to them at present.

PERU

Academic psychology. At the University of San Marcos in Lima, oldest university in the western hemisphere, psychology is scattered through several faculties. Dr. Walter Blumenfeld, a German psychologist, formerly a professor in Dresden, teaches two courses, general and experimental psychology, and is head of the Laboratory of Experimental Psychology in the Faculty of Biological Sciences. Dr. Hans Hahn, also a German psychologist and formerly professor of psychology in Heidelberg University, teaches two courses in the Faculty of Economic Sciences, one course in general and one in applied psychology. In the Faculty of Pedagogy and Letters, which prepares secondary and normal school teachers, two courses in psychology are taught by Dr. Carlos Cueto, general psychology and an advanced course which varies in theme each term. In the same faculty a course in child and adolescent psychology is taught by Manuel Argúelles. A course in judicial psychology is taught in the Faculty of Law by Dr. Honorio Delgado. Dr. Delgado is head of the Department of Psychiatry and Neuropathology of the Faculty of Medicine. He had much of his training in Germany and is probably the most well known and influential psychiatrist in South America. He has made many important contributions, both theoretical and experimental, to the psychological and psychiatric literature of South America. Dr. Delgado is co-editor with Dr. J. O. Trelles of the *Revista de Neuropsiquiatría*, one of the outstanding South American journals in this field. Two other psychiatrists assist Dr. Delgado in his university teaching in psychiatry. A one year course in psychiatry is obligatory for all medical students in their fourth year with the fifth year devoted to internship for those who go on in psychiatry. The facilities of the Victor Larco Herrera Hospital, the national mental hospital, are available for psychiatric teaching.

At the University of San Augustín of Arequipa at Arequipa in southern Peru philosophical psychology is taught in the Faculty of Philosophy and Letters, child psychology in the Faculty of Pedagogy and experimental psychology in the Faculty of Sciences. Little psychology is taught at the Catholic University in Lima or at the University of Cuzco in Cuzco, Peru.

At the National Pedagogical Institute for Girls and the National Pedagogical Institute for Boys, both in Lima, secondary and normal school teachers are trained. At each institute a course in general and a course in child psychology are taught. The institute for boys has a third course, mental measurement, under a professor who has had some training in the United States.

In the normal schools, which train primary school teachers, general psychology is taught in the second year, child development in the third year and tests and measurements in the fourth and last year. At the School of Social Service of Peru, in Lima, all students have a four semester course in psychology taught by Dr. Luis Guerra, a psychiatrist.

Research and applied psychology. Peru has a very famous research institute known as the Institute of Andean Biology (Instituto de Biología Andina) in Lima, headed by Dr. Monge. Within this organization is a Laboratory of Psychology under Dr. Hans Hahn, the German psychologist who is also at the University of San Marcos. Dr. Hahn has done a number of important and interesting research studies on the psychological effects of altitude on subjects studied at sea level and subsequently at altitudes of 15,000 feet or more, and also psychological studies of populations who live permanently at these high altitudes. Peru offers exceptional opportunities for altitude studies because many of its inhabitants live at the extreme altitudes of the Andes, far above anything to be found on the North American continent.

Dr. C. Gutierrez Noriega, Prof. of Pharmacology on the medical faculty of the University of San Marcos, is also interested in studying effects of high altitude living, but from the physiological aspects. Dr. Gutierrez Noriega has a wide range of interests which border on the psychological field.

There is an important new psychological organization in Lima which is only several years old, the Instituto Psicopedagógico Nacional, which is under the Ministry of Education but has considerable autonomy. The function of the Institute is research on problems related to education. The director is Dr. Luis Guerra, a psychiatrist who has had his professional training partly in Peru and partly in Germany. The work of the Institute is divided into six departments representing varied approaches to the study of psychology and education and one member of the staff is responsible for each of these divisions. Most of the staff have other obligations outside the Institute as well. The department of research on normal development is under Dr. Walter Blumenfeld who is also at the University of San Marcos. The research on special development is headed by Dr. Guerra himself and consists mainly of study of exceptional children. The department of anthropological research is headed by Dr. Julio C. Pretto who is working to establish anthropometric norms for Peruvian children. The department of pedagogical research is under Dr. Luis Felipe Alarco and is concerned with research on methods and materials of instruction. The department of sociology of education is the newest and is to be under Dr. Carlos Cueto who is now in the United States for study. The department of vocational guidance is under Dr. Maurice Simon, a Belgian educator who has been in Peru several years and is also the director of the National Pedagogical Institute for Boys. The work in Dr. Simon's department consists of both research and practical service. Vocational and professional guidance are available for students and others on an individual basis and employee selection projects are also undertaken. For example, the department recently undertook the selection of penitentiary guards and of taxi drivers. It is planned to use this department also to select students for the Faculty of Pedagogy of the university and for the National Pedagogical Institutes and trade schools. The Instituto Psicopedagógico Nacional is still small, with no more than the head and one or two assistants in each department, but it has made excellent progress for the length of time it has been in existence. It published its research in its own archives, *Boletin del Instituto Psicopedagógico Nacional*, of which several volumes have so far been issued.

Under the Ministry of Public Health, Labor and Social Welfare, there is a large institute concerned with the welfare of children, Instituto Nacional del Niño, headed by Dr. Manuel Salcedo. Within this organization a child guidance clinic, Clínica de la Orientación de la Niñez, has been established recently. The director of the clinic is Dr. Baltazar Caravedo, a psychiatrist, who has had some training in the United States recently. There is one other psychiatrist on the staff, two social workers and a nurse. All ages of children are accepted and with a variety of problems but most cases now referred are at the preschool age levels. Psychological tests are given by the social worker or nurse.

There is one unit for mentally disturbed, epileptic or mentally retarded children within the Victor Larco Herrera Hospital, the national mental hospital. The director of the children's work is Dr. Carlos Krumdieck, a pediatrician, who is very much interested in developing complete psychological as well as physical services for children. Dr. Luis Guerra and his staff from the Instituto Psicopedagógico Nacional do some examining and testing of children here.

There is one other applied psychology project in Lima, the laboratory of psychology in the National Police Hospital, under the direction of a pediatrician, Dr. Pastor Carranza. The function of this laboratory is the examination of all persons applying for driver's licenses, of which there are four classes, professional or private driver of an automobile, motorcycle operator and streetcar motorman. In addition to a physical examination with emphasis on neurological reactions, each candidate must pass a psychotechnical examination, consisting of a battery of psycho-motor and psycho-sensory tests. Standards are more rigorous for some types of license than for others. There is also a psychiatric unit within the police hospital, under Dr. Federico Sal y Rosas, where applicants for police work are checked for their psychiatric fitness for the work and where police already in service may receive psychiatric care when needed.

URUGUAY

Academic psychology. The University of Uruguay has little work in psychology. In each faculty a two-year sequence in philosophy is offered about half of which is devoted to psychology. So far the university has not established a faculty of humanities or philosophy in which psychology courses are customarily taught in South America. A private organization, which to some extent takes the place of such a faculty, is the Instituto de Estudios Superiores de Montevideo. The Institute has no entrance requirements nor does it grant degrees. Occasional courses in psychology are offered, usually one in contemporary psychology taught by a psychiatrist, one in experimental criminal psychology by a priest.

Psychiatry is taught by Dr. Antonio Sicco in the medical faculty of the University of Uruguay and the facilities of the Hospital de Vilardebó de Alienados, the national mental hospital, are available in the teaching of medical students. In addition to the course in psychiatry a post-graduate group has formed recently to study psychotherapy.

At the Normal Institute psychology is required in the last three years of the seven-year course, in the fifth year general psychology, in the sixth year child and adolescent psychology, in the seventh year psychopedagogy. One of the professors who teaches these courses, Rebeca Milies, is now in the United States for study.

Applied psychology. The outstanding organization for psychological service in Uruguay is the Sebastian Morey Otero Psychopedagogical Laboratory of the Normal Institutes, which comes under the primary and normal school division of the Ministry of Public Instruction. The director of the Laboratory is Maria A. Carbonell de Grompone and there is a sizable staff on which is Rebeca Milies, professor of psychology in the Normal Institute. For two years, ending recently when he went to Rio de Janeiro, Dr. Emilio Mira y López, the Spanish psychiatrist, was supervisor of a research project supported by several organizations, the purpose of which was to establish norms on the Uruguayan population for various types of psychological measurement, including Dr. Mira's own myokinetic test. The usual purpose of the Laboratory is to do psychopedagogical studies of children in the primary schools of Montevideo and of students in the normal schools, to make individual case studies of some children and to do vocational guidance. The program of the Laboratory is a combination of research and service. A bulletin of research studies, *Boletin del Laboratorio de Psicopedagogía Sebastian Morey Otero*, is issued from time to time.

A council for child welfare, Consejo del Niño, an autonomous organization under the Ministry of Public Health and Social Welfare, has among other functions a school health service and a small child guidance service. There is also one other small psychological service, a part-time psychometric service with patients at the Hospital de Vilardebó de Alienados by Margarita Marx.

Another organization in Montevideo, which is not at all psychological in function at present is, however, worthy of note because of its possible future significance in the development of mental hygiene and other psychological services. It is the International American Institute for the Protection of Children (Instituto Internacional Americano de Protección a la Infancia), to which all of the twenty one American republics, including the United States, contribute financially and to which all send delegates. A permanent staff is resident in Montevideo guided by a council made up of two delegates from each country, one local delegate from each country resident in Montevideo and one delegate resident in the country itself. The functions of the Institute are not unlike those of our Children's Bureau in Washington, except that it operates on an inter-American basis. It is a fact-finding and disseminating and a consulting organization for all aspects of child welfare in the Americas. Its three major divisions are health, education and social welfare. The emphasis to date has been mainly on health, medical problems, infant mortality and social welfare legislation affecting children. Nothing has been done on mental hygiene so far although it is hoped to extend the organization's activities into that area eventually. The Institute publishes a quarterly bulletin, *Boletin del Instituto Internacional Americano de Protección a la Infancia*, and a monthly news review. It also maintains an international library on child welfare, with complete collections from all the countries of all publications having to do with this subject. It sends out duplicate copies on loan to all parts of the continent upon request. The library has put out in pamphlet form a printed bibliography entitled, *Bibliographía de Psicología y Psicopedagogía.*

VENEZUELA

Academic psychology. Venezuela is another country in which practically no training is offered in psychology. There are no courses at either of the two universities of Venezuela, the Central University in Caracas or the University of the Andes in Mérida. Psychiatry is taught in the medical faculty of the Central University by Dr. Raúl Ramos Calles who had several years of training in the United States at such centers as Phipps Clinic and the Worchester State Hospital. Dr. Ramos Calles is also director of the Municipal Psychiatric Hospital which is the only public mental hospital in Venezuela. At the Instituto Nacional de Pedagogía in Caracas, the training center for secondary and normal school teachers, a course in general psychology is taught by a priest and a course in applied psychology, including mental tests, by a Spanish psychiatrist, Dr. José Ortega Duran.

Applied psychology. Venezuela has somewhat more to offer in the applied field than in the academic. The minister of education at the time of the writer's visit was Dr. Rafael Vegas Sánchez, a psychiatrist who, because of his own professional orientation, has been instrumental in developing several projects in which psychological services are included. The Home for the Study of Minors (Casa de Observación de Menores) receives delinquent boys through the juvenile court. Each case is studied by a social worker, a part-time psychological worker, Rosa Padlina Franzetti, and by Dr. Vegas himself. Dr. Vegas does psychotherapy with some cases. Boys who can benefit from further institutional care and training are sent on to a second small farm school, the Institución de Preorientación, near Caracas. Dr. Vegas again supervises the care there and follows a few cases psychiatrically.

In the experimental primary school, Escuela Experimental Venezuela, some psychological testing is done.

Within the Ministry of Sanitation and Social Assistance there is a division responsible for the school health program of the nation, Instituto de Higiene Escolar, whose director, Dr. José Ortega Duran, happens to be a psychiatrist. Dr. Ortega is Spanish, had both psychological and psychiatric training, was head of a psychological institute in Barcelona, and has been in Venezuela for several years. He has created a psychiatric department within the Institute for the individual study of problem children. He is assisted by another Spanish psychiatrist, Dr. M. L. Sánchez Martín, and by a psychological assistant, Anita Arraiz. Dr. Ortega is working on the standardization of various North American group tests of intelligence, personality, etc. on Venezuelan populations.

The Instituto de Higiene Mental within the Municipal Psychiatric Hospital has its own small separate building and functions as a child guidance clinic. Dr. Hernán Quijada is the head of it. The clinic is able to see about seventy five cases a year.

SUMMARY BY FIELDS OF PSYCHOLOGY

University training in psychology. The number of courses in psychology offered within a single university in South America, excluding psychiatry courses, ranges from none at all to eight. In the case of the eight, at the University of San Marcos in Lima, three are in general psychology and, therefore, by title at least, duplicates of each other and not additional courses. The most typical program in South American universities is a sequence of two or three courses, usually a first, general, largely philosophical course, a second "experimental" course and a third course which may be child and adolescent or educational psychology.

The city of São Paulo, Brazil, is the center which seems to the writer to afford the most extensive training in psychology in South America. The Faculty of Philosophy, Science, and Letters of the University of São Paulo and the Escola Livre de Sociologia e Política, if considered together, offer at least eight distinct courses, ranging from general and educational psychology to clinical and technical courses.

Even this center, however, does not have the number and variety of courses offered in a typical department of psychology in a university in this country. It can be said then that there is probably not even one university in South America where adequate training can be obtained at present for the *profession* of psychology, if that profession is thought of as independent of psychiatry and education. The obvious reason for this is that, as stated at the beginning of the article, the profession of psychology itself still does not exist as independent of other professions in South America. It is natural that when few outlets for professional work are available there should be little demand for the training that leads to that work and, therefore, a slow development of such training.

Teacher-training in psychology. Most teacher-training institutions—pedagogical institutes for the training of secondary and normal school teachers and normal schools for the training of primary school teachers—usually offer two or three courses in psychology as part of the required curriculum. These institutions in this respect compare very favorably with teacher-training institutions in the United States. It is not at all infrequent that the teacher-training institution offers more psychology than the local university. In several countries secondary schools also have a required course in philosophical psychology.

Research in psychology. South American universities usually make very little provision for *research* in psychology. Relatively few professors of psychology are personally interested in research as we know it, often being more philosophical than scientific in bent. Those who are interested usually receive very limited funds or no funds at all for specifically research purposes from the university itself and are, therefore, seriously handicapped in obtaining equipment or in employing assistants. Any research done must usually be done by the professor personally, even to the giving and scoring of tests, the collecting of experimental data and the subsequent statistical labor. Only the fortunate few have one or two part-time, often untrained assistants. One of the big sources of research in most countries, namely, masters' and doctors' theses in psychology, is entirely missing in South America because degrees are not granted in psychology.

These difficulties all limit not only the quantity but also the quality of the research done. Many of the "laboratories" of psychology in the universities are not research laboratories but teaching laboratories where students perform the classical experiments common to experimental courses everywhere. There are very few laboratories which are dedicated to research primarily.

There are a few individuals and a few organizations which are more concerned or as much concerned with research in psychology (apart from psychiatry) as with either teaching or the practical applications of psychology. Outstanding among these names are Eduardo Krapf and Horacio Rimoldi in Argentina; Decio de Souza, Noemy da Silveira Rudolfer, Manoel Lourenço Filho, Emilio Mira y López and the entire staff of the Escola Livre de Sociologia e Política de São Paulo in Brazil; Alfonso Aguirre and Abelardo Iturriaga in Chile; Mercedes Rodrigo in Colombia; Walter Blumenfeld, Honorio Delgado, Luis Guerra, Hahs Hahn, C. Gutierrez Noriega and the entire staff of the Instituto Psicopedagógico Nacional in Peru; Maria A. Carbonell de Grompone and Rebeca Milies in Uruguay and José Ortega in Venezuela.

Child guidance clinics and other psychological services for children. Few public school systems in South America as yet have programs of mental testing or other forms of scientific study of pupils for purposes of placement, guidance or analysis of individual needs or progress. In the larger cities of several countries achievement test programs are in use, mainly for purposes of determining promotions or pupil grouping, less often for the purpose of identifying pupil deficiencies. School clinics for the individual psychological study of children are beginning, as did our own years ago, with the study of mentally or academically retarded children almost exclusively and are as yet little concerned with emotional problems, behavior difficulties or personality deviations. Some of the other psychological organizations within school systems or under ministries of education are devoted to basic research needed for curriculum construction and for the development of educational materials and methods, functions which in the United States are usually filled by university departments of education.

Psychological work in children's institutions is limited with few exceptions to the institutions for delinquents or mental defectives. Institutions for dependent children or for physically or sensorially handicapped children seldom have psychological facilities of any kind.

The *concept* of a child guidance clinic—a clinic to which all types of psychological problems may be referred and in which psychotherapy is provided—has general acceptance almost everywhere in South America but so far few such clinics *exist.* Only Venezuela, Argentina, Brazil and Peru have clinics which follow the pattern of psychometric, psychiatric, physical and environmental study of each case, which accept behavior and personality problems as well as problems of mental retardation and which include provision for psychotherapy and social service therapy for at least some of the cases studied. In the field of psychological clinics associated with public school systems as well as psychological testing programs for practical purposes of pupil guidance, Brazil has progressed considerably farther than other South American countries.

Vocational guidance. Many of the South American countries have made at least a beginning in vocational guidance. In almost all of the projects of this type there is a marked tendency to rely heavily on psychological tests, in some clinics almost completely on psychological tests. Vocational guidance workers are usually strongly test-minded and give relatively less attention to evaluation of personality, interest and emotional and social factors. Brazil, again, is the country which has shown the greatest interest and activity in this aspect of applied psychology.

Industrial and personnel psychology. In this field, yet again, the name of Brazil comes up. It is indeed the only South American country where industrial psychology projects are under way on any but the smallest scale. SENAI and IDORT (see above under Brazil) and the Division of Industrial Education of the Ministry of Education and Health have the only extensive personnel and industrial programs which the writer found in South America. Brazil's greater industrial development is almost certainly the cause of its greater interest in industrial psychology.

In several countries psychological test batteries are used to aid in the selection of men for the armed forces, but again in Brazil, this policy is most extensively followed, with a well established laboratory at the aviation school and psychiatric and psychometric techniques used in the selection of other armed forces personnel.

The comparative status of other professions closely related to psychology. The following brief discussion is only for the purpose or comparing the present status of psychology with that of closely related professions and is not a discussion of those professions in and of themselves.

a. *Psychiatry.* The psychiatric profession is much farther advanced than the psychological profession in South America both as to training and as to practical acceptance. Indeed, psychology, as has already been stated, is not established at all as an independent profession. Much of the psychology taught in universities and other educational institutions is taught by psychiatrists and is inadequate at best as *professional* training. Psychiatric training on the other hand, while very poor or almost entirely lacking in some countries, in other countries reaches a standard of excellence, at least in adult psychiatry, which equals the best psychiatric centers of Europe and North America. Child psychiatry is much less well developed in either training or practice. As everywhere, one finds in South America too, less emphasis on mental hygiene or preventive aspects of psychiatric practice than on treatment of mental illness already developed.

b. *Social service.* Social service, though less advanced than psychiatry in both training and practice, is far ahead of psychology. Schools of social service in South America oddly enough are not associated with universities but exist autonomously, usually under some government ministry. The social service career is well defined and popularly accepted as an independent profession in most of the countries. Courses of training are often well worked out and, in some countries, are very complete and thorough. The demand for social workers is so great in many places that students often have positions promised to them before their training is completed. Some social service training is still found at secondary school rather than a university level but there is a trend now toward raising the standards of preparation required for admission to schools of social service and also toward lengthening the course of training. This profession offers a good example of the stimulus given to professional training by a practical demand for trained workers. Unfortunately, psychology is in a very different position.

CHIEF PROBLEMS FACING PROFESSIONAL PSYCHOLOGY IN SOUTH AMERICA

Lack of professional training. As stated earlier, there is not, in the writer's opinion, even one center in South American which offers sufficiently extensive and varied training to prepare students adequately for the career of psychology. One reason for this situation is the lack of flexibility in South American universities. They tend to continue offering mainly the classical careers, philosophy and letters, medicine, law, etc. and are slow to incorporate new profesisonal fields or new branches of specialization within old ones. It is significant that new professions often tend to develop their own training centers *outside* universities, as has happened in the case of social service, to name but one. Within the universities, too, there is a rigidity of faculty organization which does not easily permit a student to cut across the vertical lines on which curricula are constructed. He finds it difficult to organize his program around a center of interest which may require taking courses outside his own faculty, however closely related the courses may be to the field in which he wishes to specialize. Faculties often tend to be very ingrown, with a student later becoming the professor of the same two or three courses he had himself taken and continuing to cover the same ground. Such practices tend to keep the courses in psychology static and to rob them of the fresh influences always brought in by professors from elsewhere or by local professors who have gone elsewhere to study.

Lack of professional security. It is customary for professors in South America to have several positions. A university is seldom able to pay a professor a full time salary adequate to maintain him and his family. For that reason, university teaching is only one part of a professor's activities. He may have a private practice, may be on the staff of other organizations or may teach psychology or other subjects in other educational institutions such as normal schools or schools for social service, or may have other types of work even more removed from his university career. This system prevails, not so much from interest or desire on the professor's part as from economic necessity. It may have some advantages but certainly has grave disadvantages in preventing him from dedicating all his efforts to university teaching and research. He is forced to divide his interests, his time and his professional preparation.

In some of the universities, too, there is lack of security of tenure which forces professors into sharp competition with each other to maintain their positions. Considerable attention has to be given to maintaining prestige through frequent publication and through teaching in such a way as to secure their popularity with students or to avoid offending important economic, political or religious interests. These considerations hardly make for the kind of atmosphere which fosters scholarship or scientific achievement. Fortunately, the situation is more favorable in some countries.

Limited research activity. There are two important reasons for the relatively limited amount of research that is going on in psychology in South America. In the first place, research is not considered so essential or important a function of a university as it is considered in this country. Professors do not have time to do research because of their divided interests and have very limited assistance, equipment or laboratory space if they do have the time. Universities emphasize the teaching function, with research playing a relatively insignificant role. Secondly, many professors of psychology are philosophically oriented and trained and are not trained in the research techniques of science. A great proportion of the psychological publication, therefore, is philosophical rather than scientific. Though no one would deny that philosophical studies are important, too great emphasis on them tends to limit interest in experimental research of an acceptable scientific character. On the other hand, where research does go on, it has too often had a purely practical, immediate objective and is not concerned with building up a body of scientific knowledge.

Absence of employment opportunities in psychology. The difficulty if not impossibility of finding a means of livelihood through psychological studies has a marked tendency to keep students out of even the few psychology courses offered, except where the courses are required. They see no future to these studies. On the other hand, students who might take them from a broad, cultural point of view, are not attracted to them often because all too frequently they are highly abstract and philosophical rather than concerned with human behavior which we consider the primary focus of psychology. Psychology courses frequently mean to the student just more philosophy instead of introducing him to a new and fascinating field of study.

Professional isolation. This problem is perhaps the most serious of all for psychology at present in South America but it is also the one most easily capable of solution. Professional workers in psychology in South America, except for several of the countries, exist in what might almost be termed a professional vacuum which would be inconceivable in Europe or the United States. Psychological societies seldom exist within even the big cities and not at all on a nation-wide basis. There are no continental conferences in psychology nor even regional meetings of psychologists. The writer found in her travels through all of the countries of South America that the psychological workers of one country seldom knew what work was being done in psychology in even the adjoining country or who was doing it. This is not so true of psychiatry, where there is considerable contact from one country to another. In psychology there is an almost complete absence of the professional stimulation and mutual encouragement, as well as mutual criticism, arising from contacts with colleagues. Lack of contact prevents the professional unity which tends to advance the interests of a profession, elevate standards of training and practice and establish the profession on an independent basis. The lack of contact often means, too, an unfortunate duplication of effort from one country to another instead of coordinated progress into new fields of activity.

Another aspect of professional isolation is the absence of professional journals. There is not even one journal in South America which is devoted entirely to psychological material of high scientific caliber. Within each country there are psychiatric, educational and philosophical journals which occasionally publish psychological articles. There are also various psychological organizations which publish their own archives. Even those journals which do exist have small circulations. With low professional salaries, relatively few people can afford to subscribe to them. As to exchange of journals between countries the situation is even worse. Added to the economic reason again there is a lack of interest in the professional work of neighboring countries which precludes much exchange of ideas or knowledge.

Perhaps the most serious aspect of the professional isolation which still exists in South America today is the relative lack of contact with *world* literature in psychology. The reasons for this are multiple. In the first place South Americans are more apt to read French and Spanish publications than those of other countries because there is little or no language barrier involved (nearly all learn to read French in addition to their own Spanish or Portuguese) and also because the culture of France and Spain has been so dominant anyway in the history of South America. Since neither Spain nor France has been among the more productive countries in the last two decades of modern psychology, the domination of their literature has undoubtedly had a retarding effect.

In the second place, although many now read English, they have little access to United States publications. They are often unable to afford our books or journals which, in terms of their salaries, are extremely if not prohibitively expensive. Libraries are loath to build up large collections of English language books because such books serve only a limited number of readers. Very few United States books are translated into Spanish, even fewer into Portuguese and there is usually a big time lag with those that are translated. Even when individuals wish to buy our books they lack systemactic information as to all the books that are available and thence are unable to select the books most appropriate to their needs. Not uncommonly they purchase inappropriate, mediocre or out-of-date books simply because these are the only books whose titles have happened to reach their ears. The situation is very different in medicine, for eaxmple, because large numbers of South American doctors take the *Journal of the American Medical Association* and others and thus have systematic information about medical publications of this country.

PROPOSALS FOR PROFESSIONAL COOPERATION IN PSYCHOLOGY BETWEEN THE UNITED STATES AND SOUTH AMERICA

South American countries in the final analysis must solve their own professional problems. It would be useless and presumptuous for any outside nation to intervene with the idea of imposing *its* concepts as to how psychology or any other field of activity must develop in a South American country. Such intervention would rightly be resisted and resented. However, it seems to the writer that there are a number of specific ways in which the United States, either through government agencies which already exist for purposes of intellectual and cultural cooperation, or through its professional psychological organizations, might be of great assistance to South American colleagues and aid in the general development of the field of psychology in South America. The psychologists of the United States, in turn, would profit richly as well through such interchange of ideas and research.

The following proposals are directed mostly toward the two problems described above as professional isolation and lack of professional training. Some of the proposals are specific, of very limited scope and could be carried out easily. Others are more extensive, ambitious and expensive but possible of accomplishment in principle in the future, if not immediately.

Information services

1. Publication of a monthly bulletin in Spanish containing very detailed and complete digests of new professional books. The digest of each book should contain an evaluation, suggested uses for it and information regarding publisher and price. Brief digests are of little value because they do not enable a person who does not have ready access to the book for preliminary examination to know whether he should purchase it or not. The book digests contained in such publications as *The Psychological Abstracts* are too brief to be helpful and the digests of journal articles are almost useless because South Americans do not ordinarily have access to the journals themselves.

2. Prepare, maintain and keep up to date a carefully selected, classified bibliography of books, old as well as new, on every aspect of psychological and educational work. Only the most worthwhile books should be included and each should be described in some detail and uses for it suggested. Publisher and price should be given and any other pertinent information so that the South American reader will be aided in every way to select what he needs and know where to get it.

3. The same type of classified bibliography as described above should be prepared for all types of psychological tests. With tests, as with books, selection of appropriate ones for use or for study is often a hit-and-miss affair because of absence of systematic information as to *all* that is available.

4. There is great need for a bureau of professional information in psychology possibly located in Washington, to which all types of professional inquiry could be addressed. Such a bureau should be prepared to give information in regard to professional literature, professional training and all types of professional resources in the United States such as special schools, clinics, applied services, etc. South Americans now often lack even a first contact in this country to which they can direct inquiries. If such a bureau were established it should be widely publicized through announcements in all professional publications in Latin America and through announcements sent to all individuals known to be interested.

Special long-time projects

1. Sponsor a Spanish-language technical journal in psychology, preferably published in Latin America, for wide circulation in all of the Latin American countries. Such a journal should be up to the standards of typical professional journals in the United States.

2. Establish a psychological test bureau somewhere in South America for the purpose of constructing, standardizing, publishing, distributing and disseminating information about all types of psychological and educational tests. Such a bureau would coordinate the scattered, overlapping, uncoordinated efforts now being made at test construction, often by workers unqualified by training for test construction work, and would serve all Latin America. Such a bureau should be staffed at first by psychologists from this country who are specialists in the field of test construction, with assistants from Latin American countries. A careful program of test construction should be planned, developing first those types of tests most urgently needed but planning for eventual development of all types.

A further development of this project would be the establishment of several fellowships at this bureau for Latin Americans with sufficient preparation. Such Fellows would be given the opportunity to learn test construction thoroughly so that eventually they could replace the North Americans on the staff.

3. Sponsor a center for psychological studies at one of the South American universities which might be interested in developing such a center with help from the United States. Such a center might be modeled after a typical department of psychology in a North American university and its purpose would be to train students for the *profession* of psychology. The center should be staffed at first by two, three or more psychologists from the United States, representing various areas of specialization, so that a well-rounded series of courses could be offered. Students from all the Latin American countries should be eligible, with a number of fellowships offered to make it possible for more to attend. It is highly probable that, should such a center be established, many governments would offer fellowships, too, in order to build up the number of trained psychologists in their countries.

4. As a corollary to the above, post-graduate field training or internships in the various applied fields of psychology should be offered, at first probably in the United States but later, as applied work develops in Latin America, at Latin American centers, as well.

Public and professional education regarding psychology

1. Preparation of a series of articles for publication in South American medical, educational and scientific journals. These articles might include such topics, for example, as a description of a typical department of psychology in a North American university, with a list of the courses usually offered; the standards of training in psychology for the M.A. and for the Ph.D. degrees; a series of articles on the various fields

of applied psychology, including description of the work involved in each area, the training and experience required for a psychologist in each area.

2. Short, popular articles on all of the above topics for newspaper release in South America for the purpose of increasing public interest in the field of psychology.

Services which could be rendered by American psychological and educational associations to South American colleagues

1. Send announcements of annual meetings to a selected list of South Americans.

2. Invite properly qualified South Americans to become members of such organizations as the American Psychological Association, the National Education Association, the International Council on Exceptional Children, the Orthopsychiatric Association, etc.

3. Professional associations could send free introductory subscriptions of official journals to a selected list of key persons or libraries in South America, to familiarize them with these journals. Even twenty five free subscriptions, properly distributed, would be of great value.

4. Encourage psychologists in this country to include South American names in the lists of colleagues to whom they send their reprints.

Services which publishing companies might be persuaded to offer

1. Send desk copies of new professional books to key professors in South America.

2. Provide all publishing companies of professional books with a list of psychologists in Latin American countries, so that they might at least receive announcements of new books.

3. Arrange with organizations of book publishers for traveling exhibits of new professional books to go from one university or library to another in the South American countries.

4. Encourage *immediate* translation into Spanish of crucial new professional books. If possible the Spanish editions should be published in South America, where costs are lower, so as to bring such books within the financial reach of a wide range of professional readers and students who usually find it impossible to buy expensive North American publications. The most effective plan might be for the American Psychological Association to appoint a committee which would select at least one good book each year from each area of psychology to be so translated and re-published.

PARTIAL DIRECTORY OF INDIVIDUALS AND ORGANIZATIONS ENGAGED IN PSYCHOLOGICAL WORK*

ARGENTINA

Buenos Aires

Universidad Nacional de Buenos Aires:
  Facultad de Filosofía ye Letras, Instituto de Psicología
    Coriolano Alberini
    José L. Alberti (psychiatrist)
    Enrique Mouchet (retired), C. Piedras 361
  Facultad de Medicina, Cátedra de Clínica Psiquiátrica
    Juan R. Beltrán
    Gonzalo Bosch, *Director*
    Lanfranco Ciampi
    Eduardo Krapf, Maipú 1266
    Enrique Pichon-Riviere, Santa Fé 1379, 9° Piso
Escuela de Asistencia Social del Instituto de Cultura Religiosa, C. Rodriguez Peña 1054
  Dr. Luis Esteves Balado (psychiatrist)
Asociación Psicoanalítica Argentina, Juncal 655, 1° Piso
  Dr. Enrique Pichon-Rivière, *Director*
Centro de Psicología y Psiquiatría Infantil, Instituto de Pediatría, Hospital de Clínicas, Córdoba 2149
  Dra. Telma Reca de Acosta, *Director* (psychiatrist), Juncal 2101.
  Nicholas M. Tavella (psychometrics)
Centro de Psicología, Corrientes 2038
  Sra. de Aparicio, *Director*
Escuela de Adaptación, Colegio Nacional de Educación
  Dra. Carolina Tobar García (psychiatrist)
Hogar "Santa Rosa," Belgrano 2670
  Dra. Carolina Tobar García (psychiatrist)
Instituto "Sigmund Freud," Callao 257
  Bela Székely, *Director*
Instituto de Orientación Profesional del Museo Social Argentino, Viamonte 1435
  Dr. Gregorio Fingerman, *Director*
Liga Argentina de Higiene Mental:
  Servicio Social, Vieytes 489
  Instituto Neuropsiquiátrico, Giribone 1961
  Dr. Gonzalo Bosch, *Director*
Sección de Psiquiatría y Enseñanza Especial del Cuerpo Médico Escolar de la Provincia de Buenos Aires
  Dr. Néstor Vallejos Meana (psychometrics), Quintana 475, 1° Piso

Córdoba

Universidad Nacional de Córdoba:
  Facultad de Derecho y Humanidades
    Francisco W. Torres
Colegio Monserrat
  Nestor A. Pizarro

La Plata

Universidad Nacional de La Plata:
  Facultad de Humanidades y Ciencias de Educación
    Alfredo D. Calcagno
    Ernesto Figueroa
    Eugenio Pucciarelli
    F. Romero

Mendoza

Universidad de Cuyo:
  Instituto de Psicología Experimental, Facultad de Filosofía y Letras
    Horacio J. A. Rimoldi, *Director*, Corrientes 4325, Buenos Aires

Rosario

Universidad Nacional del Litoral:
  Facultad de Medicina
    Antonio Foz
    Teodoro Fracassi, Corrientes 933

---

* To facilitate direct communication with South American psychologists mailing addresses of organizations and individuals are given where necessary. When an individual's preferred mailing address is other than that of the institution under which he is listed, the address follows his name.

Santa Fé

Hospital Psiquiátrico
Dr. Luis E. Bonsembiante, *Director*, Humberto 1° 2947
Escuela de Servicio Social de Santa Fé, Calle San Martín
Dr. Luis E. Bonsembiante

BOLIVIA

La Paz

Departamento Pedagógico del Ministerio de Educación, Sección de Planes y Programas
Guido Villa-Gómez, *Director*
Elvira Pommier

Sucre

Departmento Pedagógico del Ministerio de Educación, Sección de Medidas
Dr. Alfredo Vargas, *Director*
Escuela Nacional de Maestros
Dr. Alfredo Vargas

BRAZIL

Porto Alegre

Universidade de Porto Alegre:
Faculdade de Medicina
Decio Soares de Souza, Av. Otavio Rocha, 116, 2° Andar.

Rio de Janeiro

Universidade de Brasil:
Faculdade Nacional de Filosofía, Av. Aparicio Borges
Secção de Pedagogía
Manoel Lourenço Filho (on leave)
Carlos Queiroz
Secção de Filosofía
Nilton Campos
Arthur Ramos (anthropology and psychiatry)
Jayme Grabois (psychiatrist), Av. Nilo Pecanha 155, 5° Andar
Faculdade de Medicina, Instituto de Psiquiatría, Av. Wenceslaú Braz 71
Henrique Roxo, *Director*
Januario B. Hencourt
Bennett College, Rua Marques de Abrantes 55
Heloisa Marinho, Rua Voluntarios da Patria 423
Instituto de Educação, Rua Mariz e Barros 227
Heloisa Marinho
Ministerio de Educação e Saúde
Departamento Nacional da Criança, Rui Barbosa 716
Helene Antipoff
Dr. Gustavo Lessa
Dr. Leme Lópes (psychiatrist)
Divisão do Ensino Industrial, Ministerio de Educação e Saúde, 14° Andar, Av. Graça Aranha.
Otto Schlesinger
Instituto Nacional de Estudos Pedagógicos, Ministerio de Educação e Saúde, 10° Andar.
Dr. Manoel Lourenço Filho
Servicio Nacional de Doenças Mentais
Dr. Adauto J. Botelho, *Director* (psychiatrist)
Hospital de Neuropsiquiatría Infantil, Rua Bernardo 2, Engenho de Dentro, Distrito Federal
Dr. Waldemiro Pires, *Director* (neurologist)
Marianne Wohlgemuth-Schreyer (psychologist)
Divisão da Aperfeiçoamento, Departamento Administrativo do Serviço Público, Edificio de Ministerio da Fazenda, Esplanada do Castelo
Dr. Emilio Mira y Lopez (psychiatrist)
Centro de Pesquisas Educacionais, Secretaria Geral de Educação e Cultura de Rio de Janeiro, Edificio Andorinha, Av. Almirante Barroso 81, 7° Andar.
Dr. Pedro Pernambuco Filho, *Director* (psychiatrist)
Ofélia Boisson Cardosa
Aracy Muniz Freire
Isa Goulart Macedo
Maria Jonqueira Schmidt
Departmento de Saúde Aeronautica do Exercito, Rua México, 9° Andar

Brigadeiro Médico Angelo Godinho dos Santos
Servicio Nacional de Aprendizagem Industrial (SENAI)
Divisão de Seleccão
Octavio Lins Martins, *Director*
Fundação "Getulio Vargas," Praia de Botofogo 186
J. R. R. Jubé
Clínica da Tijuca, Rua Alvarez de Brito 12
Dr. Arruda Camara, *Director* (psychiatrist)
Dra. Iracy Doyle (psychiatrist)

São Paulo

Universidade de São Paulo:
Laboratorio de Psicología, Departamento de Educação, Faculdade de Filosofía, Ciencias e Letras
Noemy da Silveira Rudolfer, *Director*
Anita Castillo M. Cabral
Clínica Psiquiátrica, Faculdade de Medicina, Av. Brig. Luis Antonio 651
A. C. Pacheco e Silva, *Dirctor*
Escola Livre de Sociología e Política de São Paulo, Largo de São Francisco 19:
Dr. Donald Pierson (sociology)
Cecilia de Castro Silva (psychology)
Dr. Robert Mange (psychotechnics)
Maria de Lourdes Campos Viegas (psychotechnics)
Dr. Duval Marcondes (psychiatrist)
Virginia Leone Bicudo (psychiatric social worker)
Dr. A. C. Pacheco e Silva (psychiatrist)
Rachel Wagner (psychology)
Clínica de Orientação Infantil da Seccão de Higiene Mental Escolar, Diretoria do Servicio de Saúde Escolar, Departamento de Educação de São Paulo, Rua Nestor, Pestana, 147
Dr. Durval Marcondes, *Director* (psychiatrist)
Dr. Joy Arrude (psychiatrist)
Dr. Mario Velez (psychiatrist)
Dr. Armando Arrude Sampaio (internist)
Maria Rita Garcoz (psychologist)
Eglantina Mondego (psychologist)
Margarida Lisboa Vieira (psychologist)
Ligia Alcantara (psychiatric social worker)
Virginia Leone Bicudo (psychiatric social worker)
Cruzado Pro-Infancia, Seccão de Psicología Infantil, Av. Brig. Luis Antonio 685
Dra. Betti Katzenstein (psychologist)
Instituto de Pesquisas Juveneis
Instituticão de Assistencia Social ao Psicopata, Rua Jaceguai 263
Dr. Paulo de Camargo, *Director* (psychiatrist)
Instituto de Organização Racional do Trabalho (IDORT), Rua de Libertade 470
Dr. Moacyr E. Alvaro, *Director* (ophthalmologist)
Dra. Aniela Meyer Ginsberg (psychologist)
Dra. Betti Katzenstein (psychologist)
Servicio Nacional de Aprendizagem Industrial (SENAI), Rua Barão Itapetininga 124, 4° Andar.
Divisão de Seleccão
Dr. Walter Barioni, *Director*
Dr. Robert Mange (psychotecnics)
Dra. Aniela Meyer Ginsberg
Dra. Betti Katzenstein
Italo Bologna
Others interested in psychology:
Dr. Cícero Cristiano de Souza (psychiatrist and psychologist)
Dr. A. Mattos Pimenta (neuro-surgeon), Marconi 138, 7° Andar
Elizabeth Forchheimer, Rua General Jardim 228, Ap. 32

CHILE

Concepción

Universidad de Concepción:
Escuela de Educación
Corina Vargas de Medina
Beatriz Coddou

Santiago

Universidad de Chile:
Facultad de Filosofía y Educación

Instituto Pedagógico, Av. Bernardo O'Higgins 2297
  Irma Salas
  Arturo Piga
  Oscar Vera, Camino Espejo 0515
Instituto Psicológico, Av. Bernardo O'Higgins 2297
  Abelardo Iturriaga J., J. D. Cañas 1859
Facultad de Medicina, Clínica de Psiquiatría, Ca. Los Olivos.
  Arturo Vivado, *Director*
  D. Ignacio Matte, Bernardo Morin 440
Universidad Católica, Av. Bernardo O'Higgins
Escuela de Servicio Social de la Junta de Beneficiencia, Augustinas 632
  Dr. Luis Muñoz (psychiatrist)
  Dra. Irma Salas (psychologist)
  Gustavo Vila
Escuela de Desarrolla, Victoria Subercaseaux 292
  Juan Sandoval
Instituto de Investigaciones Pedagógicas, Dirección General de Educación Primaria, Santa Isabel 478
  Alfonso Aguirre, *Director*
  Cecilia Weitzman de Opazo
Casa de Menores, San Francisco 1152
  Dr. Eduardo Brücher Encina (psychiatrist)
  Manuel Zamorano
  Abelardo Iturriaga J.
Politécnico "Alcivies Vicencio," San Bernardo, Chile (near Santiago)
  Dr. Ricardo Stauerwal
Ciudad del Niño "Presidente Rios," Gran Avenida 6150
  Matilde Huici
Others interested in psychology:
  Dr. Luis Tirapegui (the psychologist who made the Chilean adaptation and standardization of the 1916 Stanford-Binet scale).

Valparaiso

Dra. Amparo Arcaya (child psychiatry)

COLOMBIA

Bogotá

Universidad Nacional de Colombia:
  Sección de Psicotécnia, Facultad de Medicina, Calle 10, 13–99
    Dra. Mercedes Rodrigo, Carrera 13a, 35-66
    José García Madrid
Escuela Normal Superior, Calle 13, 16-74
  Dr. Luis Alejandro Vargas
  Dr. José Francisco Socarras (psychiatrist)
Others interested in psychology:
  Dr. Luis López de Mesa (psychiatrist)

ECUADOR

Quito

Universidad Central:
  Facultad de Ciencias Médicas
    Julio Endara
    Jorge Escudero
Instituto Superior de Pedagogía y Letras
  Edmundo Carbo
  Ermel N. Velasco
Instituto Nacional Mejía, Gabinete de Psicología
  Gerardo Larrea
Policlínica del Ministerio de Previsión Social y Trabajo
  Dr. Jorge Escudero (psychiatrist)
  Gerardo Larrea (psychometrics)
Colegio Americano
  Carlota de Garcés

PARAGUAY

Asunción

Universidad Nacional de Paraguay:
  Facultad de Humanidades
    Noemy da Silveira Rudolfer (on leave from the Universidade de São Paulo, Brazil)
  Instituto Psiquiãtrico de Asunción, Facultad de Ciencias Médicas
    Carlos Álvarez, *Director*
    Andrés Riverola, Eligio Ayala 476
Escuela Normal No. 1, General Diaz y 25 de Noviembre
  Maria Adela Garcete Speratti, Rio de Janeiro y Verde

PERU

Arequipa

- Universidad Nacional del Gran Padre San Augustín de Arequipa:

Cuzco

- Universidad Menor del Cuzco:

Lima

- Universidad Mayor de San Marcos:
  - Facultad de Ciencias Biológicas, Gabinete de Psicología Experimental, Belaochaga 556
    - Walter Blumenfeld, D. Elías 245, Miraflores, Lima
  - Facultad de Ciencias Económicas
    - Hans Hahn, Independencia 323, Miraflores, Lima
  - Facultad de Pedagogía y Letras
    - Carlos Cueto
    - Manuel Argüelles
  - Facultad de Medicina, Departamento de Psiquiatría y Neuropatología
    - Honorio Delgado, *Head of department*
    - Luis A. Guerra
- Universidad Católica
- Instituto Pedagógico Nacional de Varones
- Instituto Pedagógico Nacional de Mujeres
- Escuela de Servicio Social del Peru, Paseo Colón
  - Dr. Luis A. Guerra (psychiatrist)
- Instituto de Biología Andina, Hospital Arzobispo Loayza, Av. Alf. Ugarte
  - Dr. Hans Hahn, Independencia 323, Miraflores, Lima
- Instituto Psicopedagógico Nacional, Carabaya 1128, Edificio Hernández 308
  - Dr. Luis A. Guerra, *Director* (psychiatrist)
  - Dr. Walter Blumenfeld, D. Elías 245, Miraflores
  - Dr. Julio C. Pretto (anthropometrics)
  - Dr. Luis Felipe Alarco
  - Dr. Maurice Simon
  - Dr. Carlos Cueto
- Clínica de la Orientación de la Niñez, Av. Quemada 460 (Part of the Instituto Nacional del Niño, which is under the Ministerio de Salud Pública, Trabajo y Previsión Social)
  - Dr. Baltazar Caravedo, *Director* (psychiatrist)
  - Dr. José Mata (psychiatrist)
- Hospital Victor Larco Herrera, Avda. Ejército, Magdalena Nueva, Lima
  - Dr. Baltazar Caravedo, *Director*
  - Dr. Honorio Delgado
  - Dr. Luis A. Guerra
  - Dr. Federico Sal y Rosas
  - Dr. Carlos Krumdieck (pediatrician interested in mental hygiene of children)
- Hospital de la Policía, Av. Brasil
  - Gabinete de Psicología
    - Dr. Pastor Carranza (pediatrician)
  - Departamento de Psiquiatría
    - Dr. Federico Sal y Rosas (psychiatrist), Av. de la Colmena 758
- Others interested in psychology:
  - Dr. C. Gutierrez Noriega, Colmena 245. (Professor of Pharmacology in the Facultad de Medicina, Universidad de San Marcos.)

URUGUAY

Montevideo

- Universidad de Uruguay
  - Facultad de Medicina, Departamento de Psiquiatría, Hospital de Vilardebo, Av. Millán 2515
    - Antonio Sicco, *Director*
    - Isidro Mas de Ayala, C. Izcua Barbat 1230
    - Margarita Marx (psychometrics), Joaquin Nuñez 3035
- Instituto de Estudios Superiores de Montevideo
- Institutos Normales, C. Cuareim 1381
  - Maria A. Carbonell de Grompone
  - Zaira Gamundi
  - Magdalena Louzan
  - Rebeca Milies, Soriano 944
  - Elodia Rincon
  - América Rodriguez Frias
- Consejo del Niño
- Laboratorio de Psicopedagogía "Sebastian Morey Otero," Mercedes 1178

Maria A. Carbonell de Grompone, *Director*
Rebeca Milies, Soriano 944
Instituto Internacional Americano de Protección a la Infancia, Av. 18 de Julio, 1648
Dr. Roberto Berro, *Director* (pediatrician)

VENEZUELA

Caracas

Universidad Central:
Facultad de Medicina, Clínica Psiquiátrica
Raúl Ramos Calles, *Director*
Ricardo Álvarez
Instituto Nacional de Pedagogía
Dr. José Ortega Duran (psychiatrist and psychologist)
Hospital Municipal Psiquiátrico
Dr. Raúl Ramos Calles, *Director*
Dr. Aranda
Dr. Ricardo Álvarez
Instituto de Higiene Mental del Hospital Municipal Psiquiátrico
Dr. Hernán Quijada, *Director* (psychiatrist)
Casa de Observación de Menores
Dr. Rafael Vegas Sánchez (psychiatrist)
Rosa Padlina de Franzetti (psychometrics) Puente Yánes a Tracaborda 84
Institución de Preorientación, Los Teques (near Caracas)
Dr. Rafael Vegas Sánchez (psychiatrist)
Escuela Experimental "Venezuela"
Rosa Padlina de Franzetti
Instituto de Higiene Escolar, Av. San Martín 387
Dr. José Ortega Duran, *Director* (psychiatrist and psychologist)
Dr. M. L. Sánchez Martín (psychiatrist), Cipreses a Miracielos, No. 11
Anita Arraiz (psychometrics)

RAPAPORT, DAVID, et al. *Diagnostic psychological testing. Vol. I.* Chicago: The Year Book Publishers, Inc., 1945. Pp. xi+573.

Diagnostic Psychological Testing deals with "the theory, statistical evaluation, and diagnostic application" of a battery of tests applied to a group of 217 psychotic patients and 54 control cases seen at the Menninger Clinic. The study is reported in two volumes, of which Volume I reviewed here, reports the test results on the Wechsler-Bellevue Adult Intelligence Scale, the Babcock Deterioration Test, the Sorting Test, and the Hanfmann-Kasanian Test of Concept Formation. Volume II, reviewed by Lindner in the pages immediately following, deals with the diagnostic significance of the Rorschach Test, the Thematic Apperception Test, and the Szondi Test.

Volume I is in three parts. In part I, the *Introduction*, the author gives his opinions with regard to the use of psychological tests in a psychiatric clinic; presents the procedures and premises of the research; and outlines the plan for the presentation of the material. In part II, *Diagnostic Intelligence Testing*, are presented the results of the different tests of the Wechsler-Bellevue series and the Babcock Deterioration Test. The major part, some 275 pages, deals with various measures of *scatter* on the Wechsler-Bellevue Scale.

The phenomenon of *scatter* or variations in test performance is something which seems to have perplexed psychologists ever since batteries containing different types of tests have been employed in appraising "general intelligence." Binet was among the first to observe and describe this apparent unevenness in development when he applied the Binet-Simon Tests to groups of children. After the translation and revision of the tests by Terman and the widespread use of "mental age" scales in America, scatter on the tests, i.e., the spread over the different age levels, became a matter of interest to a number of psychologists. Doll, Pressey, Wallin and others have reported on test scatter for various groups of subjects.

In the early years of Binet testing test variation was assumed to have diagnostic significance. On the basis of scatter many claims were made which later investigations have proved untenable. Even such an astute observer as Alfred Binet claimed that scatter on the Binet-Simon Test was the "differentiating attribute of the defective child," and a number of American writers following Binet's lead have claimed that scatter on a test is symptomatic of one kind of mental disorder or another. Some have found scatter to be indicative of "potential feeble-mindedness," others have claimed that delinquents show a wider scatter than non-delinquents, still others claim that psychotic patients tend to have a wider scatter than normal patients.

As a matter of fact, wide variations in test performance have been found in each of these groups, but what these authors fail to point out is that variation in test performance is the rule rather than the exception in all test performance.

As psychologists have become better acquainted with the nature of "general intelligence," and have learned more about the theory of mental tests and test construction, they have paid less and less attention to the phenomenon of scatter as a diagnostic indicator. For a time it appeared that the arguments which had been advanced by Richardson, Lorr, Meister, and others, had been sufficient to convince most psychologists that scatter, particularly on the Binet, was as much a function of the test as it was the function of the individual tested. Now it seems that the same line of arguments must be presented all over again for the Wechsler-Bellevue. Let us hope that it will be more fruitful. This volume does not give much ground, however, for optimism in this direction.

Part III is a wordy exposition of what the author calls *Diagnostic Testing of Concept Formation.* Three types of tests are used. Some discussion is given to the Similarities Test of the Welchser-Bellevue. Most attention, however, is given to the Hanfmann-Kasanin Test of Concept Formation and the Sorting Tests of Goldstein and Scheerer used in their study of *abstract* and *concrete* behavior. Since the Hanfmann-Kasanin test does not lend itself to a quantitative approach in the same way as does the Wechsler-Bellevue, the author indulges in long qualitative comparisons of the reaction of psychotic patients on these tests. "By introducing all these tests into our battery conceptual performance can be compared with conceptual verbalization." In one of the appendices the author reports the raw scores for each of the nosological groups on each of the test—a formidable array of rather meaningless data.

There is a bibliography of 32 titles in another of the appendices and also a brief review of each of the tests used in the testing program.

The experienced clinician, upon a superficial first glance at *Diagnostic Psychological Testing*, is likely to hail it with enthusiasm. He will probably exclaim, "Here is something really important, something which will enable individual psychological testing to contribute something worth while." Upon more careful examination, however, he will be sadly disappointed. If he reads carefully and stops for a moment to analyze some of the concepts, he will find them quite meaningless. It is unfortunate that such a prodigious amount of work and effort should produce so little of real significance.

The superficial attempt to apply statistical procedures to a series of test data with apparently little understanding of the underlying theory gives the book a false appearance of objectivity. Many of the figures are almost unintelligible and one of them, for example (Fig. 2, 56), does not even agree with the text. The volume is a good illustration of what happens to test interpretation when an investigator starts out with the assumption that it is "better to present what experience teaches than to by-pass it for lack of statistical evidence" (v).

The sentence structure is poor and the exposition is almost intolerable. The following is a rather typical example of what the poor reader must unravel.

> From the study of Mean Scatter in terms of scatter ratings, we conclude, *theoretically*, that disparity of efficiency of functions underlying subtest achievements in an intelligence test is an expression of maladjustment, and that the degree of disparity—that is, the dispersal from each other and from their central tendency of these achievements and the functions underlying them—roughly parallels the degree of maladjustment; and *practically*, that great dispersal on the total test and/or on the Verbal part is diagnostic for Schizophrenia or Depressive Psychosis, while great dispersal on the Verbal part with little on a generally-lowered Performance part is specifically diagnostic of Depressive Psychosis (72). (The capitalization and punctuation is of course the author's.)

Not many readers will have either the time or the patience to ferret out the meaning of such a sentence if indeed it has any meaning, and this is typical of the exposition.

It is of course necessary and important for a psychological examiner to analyze thoroughly his test data. At the same time he must not be so carried away by his desire to make a contribution that he makes false or untenable claims. Not to make a statement or failure to be as astute as perhaps one ought to be, can, at least, do no ill, but to draw conclusions and inferences which are not at all warranted from the data (and there are many such in this volume) may lead to irreparable damage. The field of individual psychological testing is one in which conservatism is a virtue.

One could argue with a fair degree of confidence that *Diagnostic Psychological Testing* has contributed little to the improvement of clinical practice.

ANDREW W. BROWN.

University of Chicago.

RAPAPORT, DAVID, et al. *Diagnostic psychological testing, Vol. II.* Chicago: The Year Book Publishers, Inc., 1946. Pp. xi+516.

In this second volume, Rapaport and his collaborators bring to a close an ambitious project which has had as its original design a rounded survey of the chief psychometric instruments currently employed by clinical psychologists. Much credit must be given to the author and his associates for their determination to carry through to a conclusion what must often have seemed an onerous and difficult task. At the same time, just because the effort has been so great, a reader cannot help wondering whether or not the energy exerted in the fulfillment of the design was really worth the end. For, quite simply, regarded either by halves or by the whole, the product falls far short of what was intended.

The second volume is divided into five sections, all of them characterized by an unevenness of presentation which probably reflects the authors' progression in their thinking not by continued development but by flashes and eruptions of intuition. This is exemplary in literary work and makes for reader-interest, but the presentation of sober scientific data is harmed rather than helped by such a technique. When, moreover, the authors engage—as they do in this volume more than in the first—in didacticisms for which an earnest reader mainly searches evidence, this fault emerges most prominently. To lift from context the numerous instances of this kind of thing would be grossly unfair to the authors, but it is a fact that the professional reader is often left aghast at the recklessness with which hypotheses are formulated and interpretations made. Such may be forgiven to Sunday Supplement writers but not in a case where the avowed intention of a study is exhaustiveness.

The volume opens with an introductory chapter—obviously the work of Rapaport himself—which formulates the plan of the work, differentiates between the diagnostic testing of intelligence and concept formation (the material of Volume I) and the testing of personality and ideational content, discusses the concept of projection and projective procedures, and closes with an interesting, highly readable brief discourse on personality processes underlying projective reactions. Following this, the Word Association Test is dissected only to be reassembled on page 84 in what the authors believe to be a more suitable fashion for yielding not only the amount but also the kind of disturbance in a patient, and to include more words evocative of "aggression . . . affective lability, fearfulness and anal trends . . . marital relationships . . . oral conflicts . . . and bodily preoccupations."

The Rorschach Test occupies the next three hundred and some pages. There is hardly an approach to the test which does not concern the authors. Moreover, those among us who are and have been somewhat alienated from the formal societies and cults of Rorschach by their preoccupation with "scoring problems" to the exclusion of real psychological and practical understanding will find in this section some rewarding reading. Were it not for the many times the authors "take off and fly blind," guided only by the wavering beacon of their own thinking and neglecting so much the rich literature and findings of others, as well as the contributions of scientific psychology, one would readily credit them with accomplishing their aims.

Chapter IV considers the Thematic Apperception Test of Murray. Its particular field of inquiry is ideational content, reflecting as it does on motivation of behavior. To enjoy and profit from it requires the reader to accept the dynamic psychological orientation of the authors. It suffers, however, from an exclusiveness which refuses to admit the contributions of other ways and points of view from which human behavior may be regarded. Even this reviewer, whose bias is similar to Rapaport's, must regret the single-mindedness of approach herein demonstrated.

The last chapter contains comments relating to diagnosing with a battery of tests. It is essentially a brief recapitulation, but wraps and delivers the message of the work with more directness than one would have expected from their earlier diffuseness of statement.

Students will thank the authors for the inclusion of Appendix III in which a critical bibliography for the tests discussed is given. This is one of the best features of the book and will be found to be of invaluable aid in working with the material. The section by Martin Mayman particularly deserves special mention.

The statement of Dr. Rapaport in the introduction that wartime exigencies forced the presentation of this text without "statistical substantiating material" puts this matter beyond the concern of the reviewer: at the same time, it is an indisputable fact that it is in this category of appraisal that many serious objections to what is presented in Volume II can be raised.

This writer wishes, out of high personal regard for Dr. Rapaport, that he could close his review with a recommendation that the volume "belongs on the shelf of every clinical psychologist and psychiatrist." Perhaps it does—if the clinician is willing and able to discount many of the happy thoughts, sweeping independent insights and unjustifiable pronouncements which parade through the book. In sum, the field is still wide open for a sober and exhaustive survey of "the theory, statistical evaluation, and diagnostic application of a battery of tests."

ROBERT M. LINDNER.

Haarlem Lodge,
Catonsville 28, Maryland.

Alcohol, Science and Society. Twenty-nine lectures with discussions as given at the Yale Summer School of Alcohol Studies. New Haven: *Quart. J. Stud. Alcohol*, 1945. Pp. xii+475.

The purpose of this book is clearly stated in the preface.

> The 29 lectures which form the text of this book, and which will be described in the introductory lecture, are not offered as any solution to the problem of alcoholism. They are not intended as such. They are offered as the soundest, independent, nontechnical statements now available on certain medical, social, legal and religious aspects of the problem of alcoholism for the student or general reader who wishes a broad and integrated understanding of this problem. The student of sociological education may, in addition, find interest in the pedagogic experiment of which these lectures form a part.

With minor exceptions the series of lectures and discussions succeeds remarkably well in achieving this purpose. Each lecturer is distinguished in his field, and through study and investigation is especially qualified to speak regarding alcohol and alcoholism.

The outstanding feature of the book is the cooperative approach of various fields of natural and social science to a single problem normally encountered by and studied separately by each field. This treatment of a single problem from all aspects is, as Dr. Jellinek states,

> The reflection of a trend which is becoming, of late, more and more evident in all brances of science. The scientist is discovering his social responsibility, and society is discovering the utility of specialized knowledge. The scientist is coming out of his seclusion and entering into the reality of life as it is lived. . . . The realization that scientific knowledge must be put to work, that it must take part in the practical solution of social issues, has led to the creation of this School. Its aim is to make the most recent findings of science available to the general public (4).

It is not factual data from any one field of science that recommend this book but rather the integration of findings from all fields and approaches.

After two introductory lectures by Dr. Jellinke, Director of the School, there are five lectures devoted to the effects of alcohol on the individual. That alcohol is high in calories, that it is a depressant of the anaesthetic type, that there is no true toxicological tolerance developed for alcohol, that the diseases associated with chronic alcoholism result from nutritional deficiencies are facts well known to physiologists and, to a considerable extent, to men in other fields who have made any serious study of alcoholism. But certainly the implications of these facts have not been adequately recognized in other fields, and the facts themselves are either unknown or unaccepted by many intelligent laymen.

This lack of acquaintance with or resistance to scientific fact on the part of laymen is revealed in the discussions which follow the lectures. The inclusion of discussions as an integral part of the textual material is a second outstanding feature of the book. Both the lectures and the discussions were recorded and the text composed from transcriptions with a minimum of editing. The class before whom the lectures were presented was one composed of 147 students selected from many applicants, and so the members can be described as selected and intelligent laymen. Yet their questions reveal both ignorance of basic facts and a resistance to clearly established findings, the resistance apparently being based on erroneous information or upon wishful deduction. The reviewer does not feel that in making these statements he is maligning or belittling the student: if anything he is criticizing the usual scientific presentation whether in lecture or in written form. The lecturer or author makes statements which to him are clear-cut both as to explicit and implicit meaning, and unless he goes out of his way to discover students' thinking and interpretation, he will be baffled by the lack of grasp and by the misinterpretations. If baffled he can accuse his students of being ignoramuses or he can attempt to make his point clear and acceptable to the student. In this book the lecturers clearly do the latter and thus the discussions in this particular book contribute markedly to the worth of the book as a whole. They highlight specific points of the lecture, emphasize the basic viewpoint of the lecturer, and serve to give additional focus, relationships, perspective.

Following the lectures on physiological effects are five on the relation of class mores, heredity, and personality factors to alcoholism. The fact that alcoholism is centered in the upper class (4% of population) and the lower-lower class (25%) is well attested, but even in these lectures one finds some debatable handling of the hereditary question. Statistics are cited to show that 52% of inebriates had either an inebriate father or mother or both, and this is interpreted rightly as showing the risks accruing from example and neglect (environmental factors). Next statistics are presented showing that approximately 35% of chronic alcoholics have parents who are psychopathological deviates, and this is interpreted as indicating hereditary taint. Just how and when psychopathological deviation has been proved to be hereditary is not revealed in the lecture, and the reviewer seriously questions the existence of this proof. The lecturer on this subject argues rightly in all examples except psychopathological deviation that the fact that a trait or characteristic runs in families is not indicative of hereditary transmission, but in this one case he accepts hereditary transmission without comment or question.

The remaining 16 lectures are devoted in order to social factors in drinking, the effect of alcoholism on society, the control of alcoholism, and the treatment of alcoholics, the final lecture being by W. W., one of the cofounders of Alcoholics Anonymous. These chapters maintain the same objectivity and careful analysis as the first ones, but the reviewer found himself scanning rather than reading at several points. Doubtless this was due more to his own interests than to the quality and worth of the material. For example and purely as example, he is not much interested in the legal aspects, even though he will grant their importance and would concede that possibly he should be more interested. If a reader is interested in and instructed by nine-tenths of the content of a book, surely that book has merit well above the average.

In summary, this book is highly recommended to any person interested in a single aspect or in all aspects of the problem of alcoholism. It is regrettable that the paper shortage required quite such fine print.

C. N. REXROAD.

Stephens College.

MORGAN, H. K. *Industrial training and testing.* New York: McGraw-Hill, 1945. Pp. x+225.

This book is written by an engineer to present procedures in testing and training which have been empirically developed. To a psychologist the treatment is not acceptable as adequate, effective, or accurate. An appreciation of this fact can best be gained from a series of quotations. These are taken out of their context, of course, but this does not make them much less coherent or more dogmatic. These statements relate to the use of tests for selection and placement, which is discussed in the first part of the book:

> There are a number of factors that, combined, make up the working personality of a person. . . . These factors are not cryptic and do not require expert knowledge to grasp their meaning. Very simply, they are, job interest, general intelligence, social intelligence, leadership, planning, drive, follow-through. . . . All these qualities can be indicated by two simple tests taking 30 minutes (15).
>
> It is not necessary to have a trained psychologist to head such a (testing) section. It is essential that the person occupying the post shall be versed in those principles of psychology which are generally recognized as well established. These are not great in number (56, 57).

Eleven more or less standardized tests are briefly described and recommended as the most valuable. There is little recognition of the fact that a test is good or valid only insofar as it is valid for a specific purpose in a particular employment situation. Simplified norms are given for dividing applicants into five groups or levels by the tests. Subjective judgment is apparently thought to be adequate for deciding what the score or level should be for success in any particular type of work.

The organization of a training department is outlined, and specific instructional methods prescribed for shop, sales, and supervisory training. In these suggestions is incorporated much good sense based on extensive experience. Follow-up and evaluation of achievement is provided for by work review sheets which are essentially rating scales. Counseling methods of a misfit or unhappy workers are proposed, based largely on the Kuder interest test.

Evidently intended for use as a textbook, there is at the end of each chapter a summary and questions and problems. Although not acceptable for college classes, it might be used in extension or in-service training courses with considerable explanation and supplementary discussion. It would be difficult for anyone with no previous knowledge of the subject to get reliable help from the book, especially in the use of tests.

The Pennsylvania State College. BRUCE V. MOORE.

RADVANYI, LASZLO. *Public opinion measurement. A survey.* Mexico, D. F.: Instituto Cientifico de la Opinion Publica Mexicana, 1945. Pp. 88.

Has public opinion measurement attained methodological soundness; does measurement influence public opinion; should methods be supervised; how best may objectivity in findings be guaranteed; what plan of financing promotes confidence—these and other questions were presented to a selected list of persons in opinion measurement, journalism, psychology, and related fields. The answers Dr. Radvanyi presents in a report of two parts: the statistical findings, and detailed statements from representative authorities and student of public opinion. Among these the name of George Gallup is conspicuously absent.

Majority opinion considers methods now used as scientifically sound but subject to further refinement; holds that opinion measurement plays a constructive role in democratic society; holds that it may influence public opinion (though harmful, by only 18.8%); holds the belief that impartiality is best guaranteed by independent (as opposed to government) organization and financing; believes some supervision desirable (56.1%) but if done should be by scientific organizations (60.2%) and not by government (13.2%). (This latter finding is significant in view of the Congressional Committee's report of January, 1945—Reviewer).

Opinion is divided as to whether "commercial" polls harm the Institute type of opinion polls; also as to the approved arrangement for financing, the tendency being somewhat in favor of the endowed form (as in the case of NORC). The reference to "commercial" polls is obviously to the market research field, not to privately financed opinion polls.

It may be observed that though the findings are timely and significant in reflecting present viewpoints, the individuals participating represent a wide diversity of competence in the field, but with, however, most of the important ones included. The variance in views expressed lends credence to the author's belief that it constitutes a good sampling of competent opinion on opinion measurement today.

NORMAN C. MEIER.

University of Iowa.

BERNHARDT, KARL S. *Practical psychology.* New York: McGraw-Hill Book Co., 1945, Pp. x+319.

This book has grown out of lectures in University Extension at the University of Toronto and throughout the Province of Ontario. In his preface, the author states that "There seems to be a real need for a book on psychology for beginners that incorporates the accuracy and scientific viewpoint of the college text and yet is written in such a way as to be thoroughly understandable by readers who have not the benefit of the college classroom." In the judgment of the reviewer, the author has realized this objective in a most creditable fashion. The style is simple, clear, concise, and dignified. And though the book is elementary in nature, the author has not sacrificed a factual and scientific approach to its subject matter. The reviewer can happily recommend this book to any intelligent layman who desires a simple survey of the essential facts and principles of general psychology.

Organizing the material of the book under five major divisions, the author in Part I introduces the subject matter, methods, and divisions of psychology, and has a helpful chapter on efficiency in reading and study. Part II is entitled *The Origin and Control of Human Activity*, and includes a discussion of the concepts of heredity and environment and maturation, motivation, mechanisms of adjustment of frustration, influencing other people, the regulation of child behavior, feeling and emotion, and emotional control. Part III considers *Individual Differences and their Measurement*, with chapters on intelligence and intelligence tests and personality and its measurement. Social relations, vocational adjustment, and mental hygiene and sane living are the topics introduced in Part IV under the rubric of *Relations with Other People*. Part V is called *Knowing Our World*, in which cognitive processes, such as sensation, observation, learning, and thinking and imagination, are analyzed.

Each major division is introduced by a series of questions which focuses the reader's attention on the main points to be discussed. And each chapter is concluded with a schematic outline of its contents and review questions. There is also a glossary of psychological terms. While there are no footnotes or specific citations to literature, reference to standard works for further study are given at the end of the book.

This is the United States edition of the book which was first published by the Life Underwriters Association of Canada. The Canadian edition was called *Elementary Psychology*, and in the opinion of the reviewer is probably a more accurate title.

Franklin and Marshall College. PAUL L. WHITELY.

BOOKS AND MATERIALS RECEIVED

BEAUMONT, HENRY. *Psychology applied to personnel.* New York: Longmans, Green, 1946. Pp. viii+167.

BRIGHOUSE, GILBERT. *The physically handicapped worker in industry.* Bull. 13. Pasadena, Calif.: California Institute of Technology, 1946. Pp. 54.

CRAWFORD, A. B. & BURNHAM, P. S. *Forecasting college achievement. A survey of aptitude tests for higher education. Part I. General considerations in the measurement of academic promise.* New Haven: Yale Univ. Press, 1946. Pp. xxi+291.

CURRAN, DESMOND & GUTTMAN, ERIC. *Psychological medicine* (2nd Ed.) Baltimore: Williams & Wilkins, 1945. Pp. viii+246.

DEWEY, JOHN. *Problems of men.* New York: Philosophical library 1946. Pp. 424.

DUNCAN, OTIS D. & OTHERS. *Social research on health.* New York: Social Science Research Council, 1946. Pp. 212.

EDWARDS, A. L. *Statistical analysis for students in psychology and education.* New York: Rinehart & Co., 1946. Pp. xviii+360.

FLESCH, RUDOLPH. *The art of plain talk.* New York: Harper & Bros., 1946. Pp. xiii+210.

GOLDRING, WILLIAM, *et al. Experimental hypertension.* Vol. III. New York: New York Academy of Sciences, 1946. Pp. x+179.

HEWITT, E. & JENKINS, R. L. *Fundamental patterns of maladjustment. The dynamics of their origin.* State of Illinois, 1946. Pp. 110.

JOHNSON, WENDELL. *People in quandaries. The semantics of personal adjustment.* New York: Harper & Bros., 1946. Pp. xiv+532.

LINK, H. C. & HOPF, H. A. *People and books. A study of reading and book buying habits.* New York: Book Manufacturers' Institute, 1946. Pp. 166.

LOWREY, LAWSON G. *Psychiatry for social workers.* New York: Columbia Univ. Press,, 1946. Pp. iv+337.

LUCK, J. M. (Ed.). *Annual Review of physiology.* Vol. III. Stanford Univ., Calif.: Annual Reviews, Inc., 1946. Pp. viii+658.

MAIER, NORMAN R. F. *Psychology in industry.* Boston: Houghton-Mifflin, 1946. Pp. xvi+463.

MORGAN, J. J. B. *How to keep a sound mind* (Rev. Ed.). NewYork: Macmillan, 1946. Pp. vii+404.

MUNN, N. L. *Psychology. The fundamentals of human adjustment.* Boston: Houghton-Mifflin, 1946. Pp. xviii+497.

RAND, WINIFRED, SWEENY, MARY E. & VINCENT, E. LEE. *Growth* *and development of the young child* (4th Ed.) Philadelphia: W. B. Saunders, 1946. Pp. viii+481.

Sargent, W. E. *Teach yourself psychology.* Philadelphia: David McKay Co., 1946. Pp. 159.

Short, Lloyd M. *Personnel problems affecting social scientists in the National Civil Service.* Pamphlet 1. New York: Social Science Research Council, 1946. Pp. iv+35.

Symonds, P. M. *The dynamics of human adjustment.* New York: D. Appleton-Century, 1946. Pp. xiv+666.

Thorpe, L. P. *Child psychology and development.* New York: Ronald Press, 1946. Pp. xxvi+781.

Westover, F. W. *Controlled eye movements versus practice exercises in reading.* Teachers College, Columbia Univ. Contrib. to Educ., No. 917. New York: Bureau of Publications, Teachers College, Columbia Univ., 1946. Pp. viii+99.

1946 achievement testing program in independent schools and supplementary studies. Educ. Records Bull. 45. New York: Educational Records Bureau, 1946. Pp. x+63.

Theory and practice in historical study: A report of the Committee on Historiography. Bull. 54. New York: Social Science Research Council, 1946. Pp. vi+177.

No 6

EAU OF EDUCL. & PSY. RESE
LIBRARY
5150
27/2/6